

Probing unknown nonperturbative effects in $b \rightarrow s\ell\ell$ with inclusive and exclusive observables

P. Alvarez-Cartelle^{1,*}, B. Capdevila^{2,†}, E. Lunghi^{3,‡} and J. Matias^{4,§}

¹*Instituto Galego de Física de Altas Enerxías (IGFAE), Universidade de Santiago de Compostela, Santiago de Compostela, Spain and Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom*

²*Departament de Física Quàntica i Astrofísica, Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona, Martí Franquès 1, E08028 Barcelona*

³*Physics Department, Indiana University, Bloomington, Indiana 47405, USA*

⁴*Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona and Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra, Barcelona*



(Received 12 May 2026; accepted 4 September 2026; published 30 October 2026)

In this paper we revisit, from a different perspective, a long-standing question: “Is the systematic deficit observed in all branching ratios mediated by a $b \rightarrow s\mu\mu$ transition due to new physics, or to a hypothetical constant unknown universal hadronic contribution that mimics new physics?” The key observation that allows us to distinguish between these two possibilities is that nonperturbative contributions associated with $c\bar{c}$ loops affect inclusive $B \rightarrow X_s\ell\ell$ and exclusive $B \rightarrow K^{(*)}\ell\ell$ modes differently. In inclusive decays, factorizable contributions are exactly determined from data on $e^+e^- \rightarrow$ hadrons, while nonfactorizable corrections are described by resolved-photon contributions at low q^2 and by local power corrections at high q^2 . In exclusive decays, by contrast, hypothetical charming-penguin effects, beyond those already included in current uncertainty estimates, could appear, in a worst-case scenario, as a constant, universal contribution that it seems, in principle, indistinguishable from genuine new physics. We identify two observables, constructed from ratios of exclusive to inclusive modes, that can discriminate between a new physics contribution and a constant hadronic contribution. Moreover, these ratios can be measured directly by LHCb, as they do not require any normalization to $J/\psi K^{(*)}$ branching fractions from B factories. A preliminary evaluation of these observables with present data shows some preference for the new physics interpretation. In a complementary test, a comparison between inclusive measurements and the corresponding sum of exclusive modes at high q^2 similarly disfavors an explanation based on a constant hadronic contribution. Finally, we provide projections for the new observables based on expected LHCb and Belle II measurements in the near future.

DOI: [10.1103/zrp8-84fg](https://doi.org/10.1103/zrp8-84fg)

I. INTRODUCTION AND MOTIVATION

Tensions between experimental measurements and Standard Model (SM) predictions for several B -meson decays, so called, B -flavor anomalies not only remain

but have been further reinforced with very recent updates from LHCb [1] and CMS [2].

On the theory side, factorizable QCD effects are absorbed into local hadronic form factors, for which several determinations exist using both continuum approaches, most prominently light cone sum rules (LCSRs), and lattice QCD computations [3–7]. Nonlocal QCD effects are typically addressed through a combination of QCD factorization (QCdf) and phenomenological modeling of the next-to-leading-order (NLO) soft-gluon correction to the nonlocal form factor [8–10]. Existing estimates of the soft-gluon contribution either rely on computations in the light cone operator product expansion (LCOPE) at small and negative q^2 (with q^2 denoting the dilepton invariant mass squared) and on dispersive relations to continue these results to the physical region, or they adopt

*Contact author: paula.alvarez@usc.es

†Contact author: bcapdevila@icc.ub.edu

‡Contact author: elunghi@iu.edu

§Contact author: matias@ifae.es

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

phenomenological ansätze. More recently, approaches based on extended LCOPE calculations for the nonlocal form factor, combined with a z -parametrization of its analytic structure and unitarity constraints imposed via dispersion relations, have also become standard tools in the field [11,12].

Within these different theoretical frameworks, several groups perform global fits to $b \rightarrow s\ell^+\ell^-$ data [9,10,12], typically including datasets of $\mathcal{O}(200)$ observables or more. Despite differences in the treatment of hadronic effects and in the selection of experimental inputs, these analyses consistently find a tension between SM predictions and the measured data, with the quoted statistical significance varying across fits. Among the most prominent and persistent discrepancies are those observed in the central q^2 bins of the angular observable P'_5 [13,14], as well as in the branching ratio of the $B \rightarrow K\mu\mu$ decay [9,15].

Reference [8] suggested that charm rescattering, involving $D_s, D^{(*)}$ intermediate states, could generate an additional contribution to exclusive $b \rightarrow s\ell^+\ell^-$ amplitudes that enters in the same way as an effective shift in $\mathcal{C}_9$, the Wilson coefficient of the semileptonic operator

$$O_{9\ell} = \frac{e^2}{16\pi^2} (\bar{s}\gamma_\mu P_L b) (\bar{\ell}\gamma^\mu \ell). \quad (1)$$

This contribution was first estimated for the $B \rightarrow K\mu^+\mu^-$ channel in Refs. [16,17], based on heavy-hadron chiral perturbation theory (HHChPT), an extrapolation prescription to cover the full kinematic region, and phenomenological assumptions to combine the various intermediate modes. Two qualitative features emerge. First, this mechanism has difficulties to reach the size of some of the observed anomalies even under the extreme assumption of fully constructive interference among all modes. Second and more importantly, monopole and dipole photon-hadron couplings yield contributions that have opposite and same signs at low- and high- q^2 , respectively. This implies that if all contributions add constructively at low- q^2 , some cancellations are expected at high- q^2 , yielding a smaller overall effect. This observation, combined with the strong q^2 dependence of the HHChPT estimate, suggests a violation of universality that is in tension with global-fit results (including bin-by-bin analyses), which show a clear preference for a single universal contribution to $\mathcal{C}_9$.

One has to acknowledge, at this point, that the low- q^2 region in Refs. [16,17] may carry sizeable systematic uncertainties due to the extrapolation strategy from the high- q^2 estimate. Moreover, it is not yet fully settled whether this rescattering contribution is already included, in whole or in part, in the NLO calculation of Ref. [18].

On the experimental side, the branching ratios of the channels $B^{(0,+)} \rightarrow K^{(0,+)}\mu\mu$, $B^{(0,+)} \rightarrow K^{*(0,+)}\mu^+\mu^-$ and $B_s \rightarrow \phi\mu^+\mu^-$ show a consistent deficit in data with respect to SM predictions, both at the low and high- q^2 regions.¹ At high- q^2 it is customary to use a wide bin in order to smooth the impact of broad resonances. In the approach of Ref. [9], which we take as the default framework for exclusive $b \rightarrow s\ell^+\ell^-$ modes, a conservative flat uncertainty of order 10% is applied at large q^2 on top of other theory errors, in order to account for potential quark-hadron duality violations.² The observed pattern of deficits across channels, as it can be concluded from the various global fits, can be accommodated by a universal new physics (NP) contribution, which we denote by $\mathcal{C}_9^U$.

The universality of such a NP interpretation has been scrutinized in dedicated studies. In Ref. [9], the variation of $\mathcal{C}_9^U$ across different q^2 regions, bins in q^2 and across the various processes entering the global fit was tested, with results compatible with universality at the 68% confidence level. Reference [20] performed a similar analysis using a resonance-model estimate of the nonlocal form factor and found consistent conclusions, although focusing on $B \rightarrow K\ell^+\ell^-$ data only. This observed universality would be challenging to replicate with a rescattering contribution of the type advocated in Ref. [8] if the latter follows the sign pattern at low- and high- q^2 suggested by the results in Refs. [16,17].

Taken together, the previous considerations point to three logical possibilities:

- (i) The rescattering contribution is already included in the NLO calculation of Ref. [18] and should not be considered to avoid double counting. A recent analysis, see Ref. [21], demonstrated that the hadronic anomalous thresholds discussed in Ref. [8] have an exact counterpart in the corresponding quark-level diagrams. Further work along these lines might be able to clarify the extent to which partonic calculations already capture the entirety of these effects.
- (ii) The calculation of Refs. [16,17] can be trusted to describe the relative signs and the strong q^2 dependence of monopole and dipole contributions at low- and high- q^2 . Therefore this hadronic contribution is actually disfavored as a solution of the anomalies by the global fits that point to a deficit in several branching ratios and that are compatible with a flat contribution to $\mathcal{C}_9$.

¹Low- and high- q^2 denote regions below and above the two narrow charmonium resonances J/ψ and $\psi(2S)$. The precise ranges used in experimental analyses depend on the decay mode and on the adopted binning.

²In Ref. [19] these quark-hadron duality violations were estimated to be of order 3%.

- (iii) Other hypothetical unknown extra hadronic effects, possibly including the rescattering contributions of Ref. [8] but not necessarily limited to it, could conspire to generate an effective contribution that resembles a universal shift in C_9 . In the following, we denote this additional hypothetical term by $C_{9,\text{extra}}^{K^{(*)},\text{low or high}}(q^2)$ for $B \rightarrow K$ and $B \rightarrow K^*$ transitions (at low and high- q^2), respectively. Note however, that universality tests currently still lack the precision required to unambiguously resolve a nontrivial q^2 dependence. Finally, although it appears unlikely, one could contemplate issues with the experimental measurements.

While experimental and theoretical improvements are ongoing, in this paper we propose a different approach to constrain the size of a hypothetical hadronic contribution that mimics the structure of a universal NP shift in C_9 . Guided by the discussion above, we consider three possible scenarios:

- (A) Universal case. The hypothetical contribution is constant in q^2 and universal across channels. It cannot arise from the rescattering processes estimated in Refs. [16,17] because of the sign correlations between low- and high- q^2 . Instead, it can arise from some other unknown constant hadronic source, that imitates a universal NP contribution. In this scenario one can combine different channels and q^2 regions to test the universality assumption directly.
- (B) Universality among channels. The contribution is universal across channels but not constant in q^2 , for instance it could differ at low and high- q^2 . This is in tension with the fact that the deviations in branching ratios display the same sign at low and high- q^2 in several modes, including $B \rightarrow K\mu\mu$.
- (C) Nonuniversal case. The contribution is neither universal in q^2 nor universal across channels. In this case it cannot mimic the NP pattern preferred by current global fits and would typically imply an observable mode dependence and/or a pronounced q^2 dependence, neither of which is clearly supported by data at present.

Scenarios B and C are disfavored by current data and, more importantly, their viability will be tested more and more as data accumulates because they predict a nontrivial pattern across channels and kinematic regions. The only scenario that is not immediately in conflict with the observed coherence of the deviations is scenario A: a universal constant hadronic contribution that would imitate a universal NP shift. This is therefore the relevant worst-case scenario, and the one on which we focus.

The main goal of this paper is to propose a test that can rule out this worst-case possibility. Our key observation is that ratios involving inclusive and exclusive $b \rightarrow s\ell\ell$ branching fractions can discriminate between a genuine

NP contribution to C_9 and a hypothetical constant hadronic contribution: exclusive and inclusive modes are expected to display different sensitivity to charm rescattering effects. Moreover, the observables we construct do not suffer from the normalization limitations that affect absolute branching ratios, which should translate into a significantly improved experimental precision. Other scenarios, like the possible appearance of large strong phases as proposed in Ref. [22], can be likely tested with the same methodology discussed in this paper.

The use of inclusive measurements at low- and high- q^2 to confirm the NP interpretation of anomalies in the various exclusive channels is not novel. This has been done over the years (see, for instance, Refs. [23]) by focusing mostly on the low- q^2 region. More recently the analyses in Refs. [24,25] stressed that possibility of reconstructing the inclusive high- q^2 branching ratio using LHCb data only and brought in focus the importance of the high- q^2 region. In this work, we sharpen this approach but our focus is not exclusively on NP but on probing the absence of unknown hadronic contributions that mimic NP and on the construction of observables that can be measured at LHCb without need for *any* external normalizations.

The paper is organized as follows. In Sec. II we discuss the theoretical description of the exclusive $B \rightarrow K\ell\ell$ and $B \rightarrow K^*\ell\ell$ decays compared with the inclusive $B \rightarrow X_s\ell\ell$ mode, with particular emphasis on the absence of the hypothetical hadronic contribution in the inclusive case. We also discuss the structure of the Wilson coefficient C_9 in exclusive and inclusive observables. In Sec. III we define two new observables that implement our test and provide their dependence on C_9 for the relevant modes. In Sec. IV we use available data to obtain first experimental determinations of these observables in a given q^2 region. In Sec. V we explore the different behavior of the new observables as a function of $C_{9,\text{extra}}$ or C_9^{NP} and describe the possible directions of improvement in the near future. In Sec. VI we compare the inclusive computation with the sum of exclusive channels to constrain a hypothetical extra constant contribution. We conclude in Sec. VII.

II. THEORETICAL FRAMEWORK: EXCLUSIVE VERSUS INCLUSIVE

We begin with some introductory remarks on the four distinct theoretical frameworks required to describe inclusive and exclusive $b \rightarrow s\ell\ell$ decays at low and high- q^2 .

The weak effective Hamiltonian relevant to the processes considered here is [26,27]

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \sum_i C_i \mathcal{O}_i + \text{H.c.} \quad (2)$$

See Ref. [28] for explicit definitions of the relevant electromagnetic and semileptonic operators beyond

$\mathcal{O}_{9\ell}$, introduced above in Eq. (1). These other operators will not play a significant role in the present discussion, since we restrict ourselves to NP contributions to $\mathcal{C}_9$. This simplified setup is motivated by the outcome of global fits.

A. Exclusive at low and high- q^2

Exclusive decays at low- q^2 are described within the SCET-II framework. The resulting leading power amplitudes are given in terms of form factors and meson light-cone distribution amplitudes. Power corrections are nonlocal and can be formally written in terms of unknown subleading form factors and higher twist distribution amplitudes (see, for instance, the discussions in Refs. [29,30]). This lack of knowledge of the nonperturbative inputs required to estimate these nonlocal effects (commonly referred to as “charming penguins”) yields uncertainties whose size is the subject of an animated debate in the literature.

In line with the analyses of Refs. [9,28,31,32], we adopt the phenomenological model of Ref. [3] and include a q^2 , polarization and process-dependent nonlocal contribution to $\mathcal{C}_{9\ell}^{\text{eff}}$

$$\mathcal{C}_{9\ell}^{\text{eff}} \rightarrow \mathcal{C}_{9\ell}^{\text{eff}} = \mathcal{C}_{9\ell,\text{pert}}^{\text{SM}} + \mathcal{C}_{9\ell}^{\text{NP}} + \mathcal{C}_{9,j}^{c\bar{c}B \rightarrow K^{(*)}}, \quad (3)$$

where $j = \perp, \parallel, 0$ and

$$\mathcal{C}_{9\ell,\text{pert}}^{\text{SM}} = \mathcal{C}_{9\ell}^{\text{SM}} + Y(q^2), \quad (4)$$

where the definition of $Y(q^2)$ can be found in Ref. [33], while the corresponding α_s corrections are given in Ref. [18]. Finally, $\mathcal{C}_{9,j}^{c\bar{c}B \rightarrow K^{(*)}} = s_j \mathcal{C}_{9,j}^{\text{KMPW}} c\bar{c}B \rightarrow K^{(*)}$ denotes the phenomenological model for the soft-gluon nonfactorizable contribution computed in Ref. [3] and that is included in the SM predictions of Refs. [9,28,31,32]. In the $B \rightarrow K^* \ell \ell$ case, this contribution depends on the polarization of the underlying transversity amplitude denoted by j . Since the analysis of Ref. [3] did not include the imaginary part of the corresponding nonlocal form factor, the parameters s_j were introduced in Ref. [31] and allowed to vary in the interval $[-1, 1]$ in order to account for the uncertainty associated with the relative phase between the short-distance and long-distance contributions to $\mathcal{C}_9$.

In recent years, significant theoretical progress has been made in evaluating the long-distance charm contribution entering $\mathcal{C}_9^{\text{eff}}$, either through explicit computations or by using resonance data to model it [34]. More recently, Ref. [35] presented a more complete calculation, including the imaginary part and higher-twist corrections. That analysis confirmed the qualitative picture obtained in Ref. [3], while improving the precision through updated inputs and a more refined treatment. Importantly, Ref. [35] found that the numerical impact of the soft-gluon contribution to the nonlocal form factor is reduced when including all light-cone distribution amplitudes. This result is an important milestone in quantifying the impact of charm contributions

in this decay. By contrast, this contribution is not included in $B \rightarrow K \mu^+ \mu^-$ decays, since both Refs. [3,35] find it to be numerically negligible for this channel.

According to the decomposition in Eq. (3), any additional hypothetical hadronic contribution, arising for instance from higher-order nonperturbative charm effects or from some other unknown constant hadronic source, would enter as an extra term on top of $\mathcal{C}_9^{\text{eff}}(q^2)$ in Eq. (3). Such a contribution would therefore compete with the long-distance charm component already included in our analysis, for which we adopt a deliberately conservative treatment. This additional hypothetical term corresponds to what we denoted in the introduction as $\mathcal{C}_{9,\text{extra}}^{K^{(*)},\text{low}}(q^2)$ for $B \rightarrow K$ and $B \rightarrow K^*$ transitions (at low- q^2), respectively. Since in this work we focus on the worst-case scenario A, namely a constant hadronic contribution capable of mimicking NP, we will mainly focus our attention on the case in which this term is process, helicity and q^2 independent,

$$\mathcal{C}_{9,\text{extra}}^{K^{(*)},\text{high}} = \mathcal{C}_{9,\text{extra}}^{K^{(*)},\text{low}} = \mathcal{C}_{9,\text{extra}}. \quad (5)$$

For the local form factors, at low- q^2 we use LCSR only inputs. At high- q^2 we rely on lattice-QCD determinations, treating the two kinematic regions independently through their corresponding z -parametrizations.

Exclusive decays at high- q^2 are usually discussed in the context of an expansion in $1/q^2$ [36]. The resulting amplitude is thus expressed in terms of form factors and local power corrections only. If this large q^2 expansion is interpreted as an OPE, q^μ is the conjugated momentum associated with the spacetime separation between the insertion of the operators $\mathcal{O}_{1,2}$ and the electromagnetic current (which produces the two leptons); thus, the charmonium resonances are precisely the intermediate states which are included via quark-hadron duality. This shows that the Krüger-Sehgal (KS) approach of Ref. [37] in this situation is meaningless (it would be akin to trying to model resonances in $e^+ e^- \rightarrow \text{hadrons}$ while having at hand the OPE). In the low-recoil region, the corresponding $\mathcal{C}_9^{\text{eff}}$ is given by [38]

$$\begin{aligned} \mathcal{C}_{9\ell}^{\text{eff}}(q^2) = & \mathcal{C}_{9\ell}^{\text{SM}} + \mathcal{C}_{9\ell}^{\text{NP}} + h(0, q^2) \left[\frac{4}{3} \mathcal{C}_1 + \mathcal{C}_2 + \frac{11}{2} \mathcal{C}_3 \right. \\ & \left. - \frac{2}{3} \mathcal{C}_4 + 52 \mathcal{C}_5 - \frac{32}{3} \mathcal{C}_6 \right] \\ & - \frac{1}{2} h(m_b, q^2) \left[7 \mathcal{C}_3 + \frac{4}{3} \mathcal{C}_4 + 76 \mathcal{C}_5 + \frac{64}{3} \mathcal{C}_6 \right] \\ & + \frac{4}{3} \left[\mathcal{C}_3 + \frac{16}{3} \mathcal{C}_5 + \frac{16}{9} \mathcal{C}_6 \right] \\ & + 8 \frac{m_c^2}{q^2} \left[\frac{4}{9} \mathcal{C}_1 + \frac{1}{3} \mathcal{C}_2 + 2 \mathcal{C}_3 + 20 \mathcal{C}_5 \right], \quad (6) \end{aligned}$$

where 4-quark operators are included. This expression receives α_s corrections that can be found in Ref. [38]. See the same reference for the definition of $h(m_q, q^2)$. Again here in addition one would have to consider any hypothetical extra hadronic contribution denoted by $\mathcal{C}_{9,\text{extra}}^{K^{(*)},\text{high}}(q^2)$.

This completes all the possible SM contributions, but of course, one also has the NP contribution to the Wilson coefficient, $\mathcal{C}_{9\ell}^{\text{NP}}$ in Eq. (3) and Eq. (6). From the global analysis of more than 200 observables in Ref. [9], there is a clear preference for a lepton-flavor universal NP contribution, i.e., $\mathcal{C}_{9\mu}^{\text{NP}} = \mathcal{C}_{9e}^{\text{NP}} = \mathcal{C}_9^{\text{NP}}$.

For the local form factors entering the high- q^2 exclusive amplitudes, we use lattice-QCD inputs fitted through an appropriate z -expansion. No combined fit with the low- q^2 determinations is performed.

B. Inclusive at low and high- q^2

Inclusive decays at low- q^2 are described in terms of an OPE. Note that the momentum conjugated to the coordinate separation between the insertion of any two operators is the four-momentum of the X_s state. In this sense, the sum over all hadronic states with one unit of strangeness is described, modulo quark-hadron duality violating effects, by the OPE itself. This picture fails when the operators $\mathcal{O}_{1,2}$, which contain a pair of charm fields, are considered. The requirement to include only cuts through the two leptons, while neglecting purely hadronic ones, exposes the OPE to contributions which diverge in the limit in which the widths of the narrow resonances vanish [39]. These resonant effects stem from diagrams which, at the quark level, involve a charm loop.

In order to address this issue, it is convenient to separate factorizable and nonfactorizable charm loop diagrams. The former can be exactly described in terms of the $e^+e^- \rightarrow$ hadrons data using a dispersion relation (KS mechanism). Intuitively, these factorizable effects yield an OPE which is local in the X_s separation but nonlocal in the separation between the insertion of the two operators and the emission of the virtual photon. The replacement of the perturbative two loop charm contribution to the OPE with the KS ansatz yields a mere 2% shift in the integrated branching ratio at low- q^2 [23].

Nonfactorizable diagrams can in principle be described by the $O(1/m_c^2)$ power corrections first discussed in Ref. [40]. Unfortunately, as was discovered in an SCET-I analysis of inclusive modes at low- q^2 , the virtual-photon emission from the charm loop introduces a dependence on the hard-collinear scale in the anticollinear (with respect to X_s) direction. The resulting contributions to the SCET-I OPE appear in the subleading order and survive after integration over the low- q^2 region. These corrections are commonly referred to as

resolved photon contributions and, while currently not known, have been estimated to be $O(5\%)$ [41–43].

Inclusive decays at high- q^2 admit the same OPE as at low- q^2 with few crucial differences. First, the OPE is an expansion in $1/(m_b - \sqrt{q^2})$ and breaks down at the end point of the spectrum. This problem appears explicitly in the expressions for the power corrections which have integrable singularities at $q^2 = m_b^2$. Focusing on the high- q^2 integrated rate, the OPE breakdown results in large power corrections which completely dominate the theoretical uncertainty on the SM prediction. The solution to this problem is to consider the high- q^2 integrated branching ratio normalized to the $B \rightarrow X_u \ell \nu$ branching ratio measured with the same q^2 cut [44]. The OPE breakdown in both numerator and denominator is almost identical and cancels in the ratio. The resulting observable has a power expansion which is under fairly good control [23,25].

The second major difference with respect to the inclusive at low- q^2 is that for $q^2 > m_{\psi(2S)}^2$ we still need to integrate over four broad resonances (rather than having to deal with the tail of the J/ψ only). As at low- q^2 , the factorizable contributions to these resonances are described using the KS method while nonfactorizable effects on the integrated rate are given by $O(1/m_c^2)$ which, in this case, are local and given in terms of standard heavy-quark effective theory matrix elements [40]. The presence of these broad resonances inside the integration region leads to fairly large effects: the replacement of the perturbative two loop charm contribution to the OPE with the KS ansatz yields a shift of about -10% of the integrated branching ratio [23]. The extent to which nonfactorizable effects in the production of the broad charmonium resonances above the $\psi(2S)$ are captured by local $O(1/m_c^2)$ corrections is currently under investigation.

III. DEFINITION OF OBSERVABLES AND NP SENSITIVITY

In this section we present two observables designed to discriminate between the two possibilities under scenario A: a universal NP contribution $\mathcal{C}_9^{\text{NP}}$ or a hypothetical constant hadronic contribution $\mathcal{C}_{9,\text{extra}}$. Notice, however, that for the analysis of each of the observables in this section we will not yet enforce universality, so we will keep each contribution $\mathcal{C}_{9,\text{extra}}^{K^{(*)},\text{high or low}}$ as formally independent.

The observables proposed in this paper are constructed from combinations of isospin averaged branching ratios and are defined as follows:

$$\mathcal{F}_{K^{(*)}}^{[1.1,6]} = \frac{\mathcal{B}(B \rightarrow K^{(*)} \mu \mu)_{[1.1,6]}}{\mathcal{B}(B \rightarrow X_s \ell \ell)_{[1,6]}}, \quad (7)$$

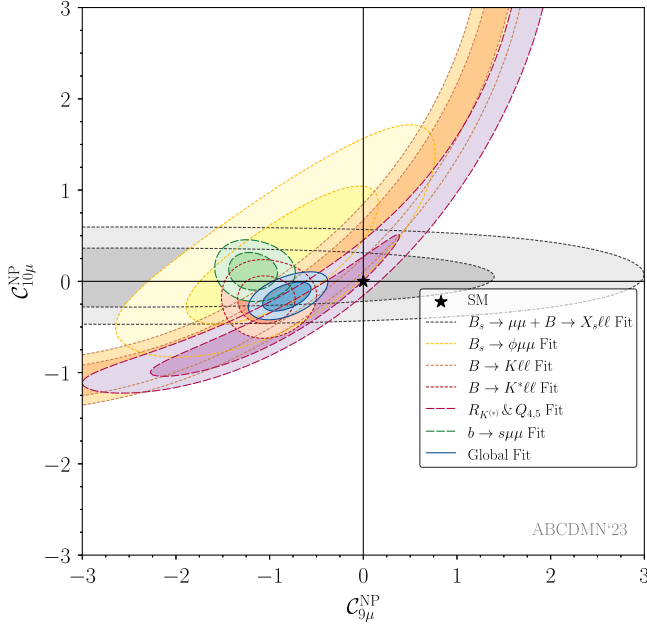


FIG. 1. Global fit taken from Ref. [9], 1σ (dark-shaded) and 2σ (light-shaded) confidence regions for $(C_{9\mu}^{NP}, C_{10\mu}^{NP})$ scenarios. Distinct fits are performed separating each of the $b \rightarrow s\ell^+\ell^-$ by modes (short-dashed contours), the lepton-flavour universality violation observables and the combined $b \rightarrow s\mu^+\mu^-$ modes (long-dashed contours), and the global fit (solid contours).

$$\mathcal{F}_{K^{(*)}}^{>15} = \frac{\mathcal{B}(B \rightarrow K^{(*)}\mu\mu)_{>15}}{\mathcal{B}(B \rightarrow X_s\ell\ell)_{>15}}, \quad (8)$$

for $q^2 \in [1.1, 6] \text{ GeV}^2$ and $q^2 > 15 \text{ GeV}^2$, respectively.³

The key idea behind the construction of these four quantities is that only exclusive modes are affected by hypothetically large nonlocal power corrections, which we indicate by $C_{9,\text{extra}}^{K^{(*)},\text{high or low}}$, whereas both exclusive and inclusive modes are affected by the universal NP contribution.

For completeness in Fig. 1 we reproduce the confidence regions in the $(C_{9\mu}^{NP}, C_{10\mu}^{NP})$ plane from Ref. [9]. These show a clear preference for a nonzero C_9^{NP} contribution, and excellent compatibility between the global fit and inclusive $B \rightarrow X_s\ell\ell$ decays, despite the uncertainty in the determination from the inclusive modes being large.

We work under the assumption that the only uncontrolled theoretical uncertainty is the size of a hypothetical hadronic constant contribution to the exclusive channels which we parametrize as a constant shift to the Wilson coefficient of the O_9 operator: $C_9 \rightarrow C_9 + C_{9,\text{extra}}^{K^{(*)},\text{low or high}}$. In this scenario numerators and denominators of the observables in Eqs. (7) and (8) are second order polynomials in $C_9^{NP} + C_{9,\text{extra}}^{K^{(*)},\text{low or high}}$ and C_9^{NP} , respectively.

³Note that for the inclusive at low- q^2 we use the more conventional $[1, 6] \text{ GeV}^2$ window for which all existing theoretical and experimental studies have been performed.

It is clear that independent measurements of inclusive and exclusive modes have the ability to independently determine C_9^{NP} and $C_{9,\text{extra}}^{K^{(*)},\text{low or high}}$. This, for instance, can be seen from Fig. 5 of Ref. [25] where it is shown that inclusive modes can confirm the presence of a nonvanishing C_9^{NP} (failure to observe a tension in inclusive modes would suggest a resolution of the exclusive anomalies in terms of $C_{9,\text{extra}}^{K^{(*)},\text{low or high}} \neq 0$). See also the discussion in Sec. IV.

In this letter we argue that the observables $\mathcal{F}_{K^{(*)}}^{[1.1,6]}$ and $\mathcal{F}_{K^{(*)}}^{>15}$ have a reduced sensitivity to C_9^{NP} given the partial cancellation between numerator and denominator and this property can be used to identify the presence or not of a non-negligible $C_{9,\text{extra}}^{K^{(*)},\text{low or high}}$ contribution to the exclusive mode.

The ingredients we need for this analysis are the full error budgets and NP formulas for the various inclusive and exclusive branching ratios.

The relevant formulas for the inclusive modes are [23]

$$\mathcal{B}(B \rightarrow X_s\ell\ell)_{[1,6]} = (17.44 + 3.31C_9^{NP} + 0.56C_9^{NP2}) \times 10^{-7}, \quad (9)$$

$$\begin{aligned} \mathcal{R}(B \rightarrow X_s\ell\ell)_{>15} &= \frac{\mathcal{B}(B \rightarrow X_s\ell\ell)_{>15}}{\mathcal{B}(B \rightarrow X_u\ell\nu)_{>15}} \\ &= (27.00 + 6.32C_9^{NP} + 0.93C_9^{NP2}) \times 10^{-4}. \end{aligned} \quad (10)$$

We assume that the uncertainties on these two observables are identical to those on the SM predictions $\delta\mathcal{B}(B \rightarrow X_s\ell\ell)_{[1,6]} = 7.4\%$ and $\delta\mathcal{R}(B \rightarrow X_s\ell\ell)_{>15} = 7.2\%$.

The low- q^2 branching ratio is the average of electron and muon channel, includes QED corrections and corresponds to what B-factories measure. The high- q^2 ratio, on the other hand, is optimized for LHCb and, thus, does not include any QED effect. As we discussed in Sec. II, only the ratio $\mathcal{R}$ can be calculated reliably in terms of an OPE. The theoretical prediction for the high- q^2 branching ratio, required to construct the observables $\mathcal{F}_{K^{(*)}}^{>15}$, can be obtained by combining Eq. (10) with the experimental measurement of the semileptonic branching ratio at high- q^2 , under the assumption of the absence of NP contributions to the latter. Using the $B \rightarrow X_u\ell\nu$ spectrum [45], the high- q^2 $B \rightarrow X_u\ell\nu$ branching ratio is determined to be [25]

$$\mathcal{B}(B \rightarrow X_u\ell\nu)_{>15}^{\text{exp}} = (1.52 \pm 0.28) \times 10^{-4}. \quad (11)$$

The high- q^2 branching ratio in the presence of NP contributions is, therefore

$$\mathcal{B}(B \rightarrow X_s \ell \ell)_{>15} = (4.10 + 0.96\mathcal{C}_9^{\text{NP}} + 0.14\mathcal{C}_9^{\text{NP}2}) \times 10^{-7}, \quad (12)$$

with a total uncertainty $\delta\mathcal{B}(B \rightarrow X_s \ell \ell)_{>15} = \sqrt{(7.2\%)^2 + (18.4\%)^2} = 19.8\%$.

The branching ratios of the exclusive modes can be expressed as quadratic functions of NP contribution to $\mathcal{C}_9^{\text{NP}}$ and of a residual hadronic contribution, $\mathcal{C}_{9,\text{extra}}^{K^{(*)},\text{low or high}}$, which (if nonzero) is not accounted for by the calculation of the nonlocal form factor at LO and our models for its NLO contributions. In scenario A we take these residual hadronic contributions to be constant and equal to $\mathcal{C}_{9,\text{extra}}$. Explicit expressions for the $K^{(*)}$ branching ratios at low- and high- q^2 , following the approach in Ref. [9], are collected in Tables I and II.

The four observables $\mathcal{F}_{K^{(*)}}^{[1,6],>15}$ can now be easily constructed combining the expressions of the inclusive branching ratios given in Eqs. (9) and (12) as well as the corresponding expressions for the exclusive modes given in Tables I and II. Note that the four $\mathcal{F}_{K^{(*)}}^{[1,6],>15}$ observables are given by ratios whose numerators and denominators are quadratic polynomials in $\mathcal{C}_9^{\text{NP}} + \mathcal{C}_{9,\text{extra}}^{K^{(*)},\text{low or high}}$ and $\mathcal{C}_9^{\text{NP}}$, respectively.

As we discuss below, these observables may offer a unique way to disentangle $\mathcal{C}_9^{\text{NP}}$ from $\mathcal{C}_{9,\text{extra}}^{K^{(*)}}$ at low- and high- q^2 . Moreover, since they involve only ratios of partial widths, they can be measured entirely by LHCb without requiring external measurements of normalizing modes (e.g., $B \rightarrow J/\psi K^{(*)}$) from BABAR or Belle.

TABLE I. Theoretical predictions for $10^7 \times \mathcal{B}(B \rightarrow K^{(*)}\mu^+\mu^-)$ written in an approximate quadratic form $\alpha + \beta\tilde{\mathcal{C}}_9 + \gamma\tilde{\mathcal{C}}_9^2$, where $\tilde{\mathcal{C}}_9 = \mathcal{C}_9^{\text{NP}} + \mathcal{C}_{9,\text{extra}}^{K^{(*)},\text{low or high}}$.

	Region (q^2)	α	β	γ
$10^7 \times \mathcal{B}(B \rightarrow K\mu^+\mu^-)$	[1.1, 6]	1.727	0.399	0.051
$10^7 \times \mathcal{B}(B \rightarrow K\mu^+\mu^-)$	>15	1.105	0.252	0.033
$10^7 \times \mathcal{B}(B \rightarrow K^*\mu^+\mu^-)$	[1.1, 6]	1.931	0.342	0.064
$10^7 \times \mathcal{B}(B \rightarrow K^*\mu^+\mu^-)$	>15	2.640	0.600	0.082

TABLE II. Uncertainties $10^7 \times \delta\mathcal{B}(B \rightarrow K^{(*)}\mu^+\mu^-)$ on the theoretical predictions for the various exclusive modes written in an approximate quadratic form $\alpha + \beta\tilde{\mathcal{C}}_9 + \gamma\tilde{\mathcal{C}}_9^2$, where $\tilde{\mathcal{C}}_9 = \mathcal{C}_9^{\text{NP}} + \mathcal{C}_{9,\text{extra}}^{K^{(*)},\text{low or high}}$.

	Region (q^2)	α	β	γ
$10^7 \times \delta\mathcal{B}(B \rightarrow K\mu^+\mu^-)$	[1.1, 6]	0.133	0.032	0.003
$10^7 \times \delta\mathcal{B}(B \rightarrow K\mu^+\mu^-)$	>15	0.147	0.034	0.003
$10^7 \times \delta\mathcal{B}(B \rightarrow K^*\mu^+\mu^-)$	[1.1, 6]	1.036	0.177	0.035
$10^7 \times \delta\mathcal{B}(B \rightarrow K^*\mu^+\mu^-)$	>15	0.235	0.056	0.005

IV. EXPERIMENTAL DETERMINATIONS OF $\mathcal{F}_{K^{(*)}}^{[1,6]}$ AND $\mathcal{F}_{K^{(*)}}^{>15}$

In this section we show how to combine existing experimental measurements of various inclusive and exclusive modes to obtain the observables $\mathcal{F}_{K^{(*)}}^{[1,6]}$ and $\mathcal{F}_{K^{(*)}}^{>15}$.

LHCb can only measure normalized ratios of the various exclusive $b \rightarrow s\mu\mu$ modes. For instance,

$$\mathcal{B}(B^+ \rightarrow K^+\mu^+\mu^-) = \left[\frac{\Gamma(B^+ \rightarrow K^+\mu^+\mu^-)}{\Gamma(B^+ \rightarrow K^+J/\psi)} \right]_{\text{LHCb}} \times [\mathcal{B}(B^+ \rightarrow K^+J/\psi)]_{\text{external}}. \quad (13)$$

In order to propagate correctly uncertainties stemming from the charmonium normalizations, we first determine the experimental averages for the two factors in Eq. (13) (similar decompositions hold for all the other modes) and then combine them. Averages for the normalizing charmonium modes are presented in Appendix A. The combinations of the LHCb and CMS measurements at low- and high- q^2 for the normalized ratios [i.e., the first factor in Eq. (13)], are presented in Appendixes B 1 and C 1, respectively. The averages of the low- and high- q^2 $b \rightarrow s\mu\mu$ branching ratios, obtained from the product of these two terms, are presented in Appendixes B 2 and C 2.

For the construction of the various $\mathcal{F}$ ratios we need also averages for the inclusive $B \rightarrow X_s \ell^+ \ell^-$ measurements. At low- q^2 we calculate the weighted average of the BABAR [46,47], Belle [48,49] and Belle II [50] measurements of the inclusive branching ratio for $q^2 \in [1, 6]$ GeV²:

$$\mathcal{B}(B \rightarrow X_s \ell \ell)_{[1,6]} = (1.44 \pm 0.21) \times 10^{-6}, \quad (14)$$

where the electron and muon modes have been averaged. At high- q^2 we employ a completely different strategy which is based on the observation that the inclusive is essentially saturated by summing over the K , K^* , $(K\pi)_{\text{S-wave}}$ and $K\pi\pi$ modes, all of which have been already measured at LHCb. The inclusive isospin-averaged high- q^2 branching ratio as sum of exclusive can be written as

$$\begin{aligned} \mathcal{B}(B \rightarrow X_s \mu\mu)_{>15} &= \mathcal{B}(B \rightarrow K\mu\mu) + \mathcal{B}(B \rightarrow K^*\mu\mu) \\ &\quad + \mathcal{B}(B \rightarrow (K\pi)_{\text{S-wave}}\mu\mu) + \mathcal{B}(B \rightarrow K\pi\pi\mu\mu) \\ &= (2.80 \pm 0.12) \times 10^{-7}, \end{aligned} \quad (15)$$

where the various contributions are given in Eqs. (C22), (C25), (C29), and (C32).

The $\mathcal{F}_{K^{(*)}}^{[1,6]}$ ratios can only be constructed by constructing separately the exclusive and inclusive branching ratios because the low- q^2 inclusive branching ratio is currently only accessible at B-factories. Using the isospin averages in Eqs. (B23) and (B24) we immediately obtain

$$\mathcal{F}_K^{[1.1,6]} = 0.084 \pm 0.013, \quad (16)$$

$$\mathcal{F}_{K^*}^{[1.1,6]} = 0.110 \pm 0.017. \quad (17)$$

At high- q^2 , employing the decomposition of the inclusive branching ratio given in Eq. (15), we can write

$$\mathcal{F}_K^{>15} = \left[1 + \frac{\mathcal{B}(K^* \mu \mu) + \mathcal{B}((K\pi)_S \mu \mu) + \mathcal{B}(B \rightarrow K\pi\pi\mu\mu)}{\mathcal{B}(K\mu\mu)} \right]^{-1} = 0.309 \pm 0.015, \quad (18)$$

$$\mathcal{F}_{K^*}^{>15} = \left[1 + \frac{\mathcal{B}(K\mu\mu) + \mathcal{B}((K\pi)_S \mu \mu) + \mathcal{B}(B \rightarrow K\pi\pi\mu\mu)}{\mathcal{B}(K^* \mu \mu)} \right]^{-1} = 0.629 \pm 0.021, \quad (19)$$

where the various ratios of branching ratios are computed using the results in Eqs. (C22), (C25), (C29), and (C32), including the impact of correlations that stem from common normalizations.

V. CURRENT STATUS AND FUTURE PROJECTIONS

We are now ready to show how to use the four $\mathcal{F}$ observables that we introduced in Eqs. (7) and (8) to discriminate between the interpretation of the anomalies in terms of $\mathcal{C}_9^{\text{NP}}$ or hypothetical $\mathcal{C}_{9,\text{extra}}^{K^{(*)},\text{low or high}}$ contributions. Notice that the analysis that we will present here is done observable by observable and it allows to formally test the corresponding $\mathcal{C}_{9,\text{extra}}^{K^{(*)},\text{low or high}}$ associated with each observable, so we keep this general notation in the plots. However, the preference of global fits, as well as of the bin-by-bin and mode-by-mode analysis done in Ref. [9] (see also Ref. [20]) for a universal constant contribution coming from either NP or from a hypothetical constant hadronic contribution, imposes universality. Therefore, the results of the global fits can be interpreted as a non vanishing $\mathcal{C}_9^{\text{NP}}$ or a universal $\mathcal{C}_{9,\text{extra}}$ equal to -1.16 ± 0.17 [9]. In order to discriminate between the two possibilities we compare the measured value of the four $\mathcal{F}$ ratios, given in Eqs. (16), (17), (18), and (19), to the theoretical predictions obtained from the formulas given in Sec. III under the assumptions

$(\mathcal{C}_9^{\text{NP}}, \mathcal{C}_{9,\text{extra}}) = (-1.16, 0)$ or $(0, -1.16)$. The results are summarized in Table III. In particular, the two high- q^2 ratios show a slight preference for $\mathcal{C}_9^{\text{NP}}$ at 1.9σ (1.2σ) for the K (K^*) final state; the two low- q^2 ratios are compatible with both scenarios at less than 1σ . Note that the uncertainties on the theoretical predictions for different values of $\mathcal{C}_9^{\text{NP}}$ and $\mathcal{C}_{9,\text{extra}}$ are 100% correlated.

In Fig. 2 we decouple the analysis presented in Table III from the actual results of the global fits. We consider all four $\mathcal{F}$ ratios and for each we show the current experimental determination (dashed green lines) and the corresponding theoretical predictions for the $\mathcal{C}_9^{\text{NP}} \neq 0$ and $\mathcal{C}_{9,\text{extra}} \neq 0$ scenarios (blue and red lines, respectively); the green, blue and hashed bands correspond to the allowed 68% confidence level (CL) ranges. In each plot the gray bands display the 1σ , 2σ and 3σ ranges allowed by the global fit.

The discriminating power of each $\mathcal{F}$ ratio depends crucially on the uncertainties on the two theoretical predictions (for $\mathcal{C}_9^{\text{NP}} \neq 0$ and $\mathcal{C}_{9,\text{extra}} \neq 0$) and on its experimental determination. In the balance of this section we discuss future prospects based on the current recorded luminosity at LHCb and Belle II and their future expectations.

Let us begin with a discussion of the experimental measurements of these ratios.

At low- q^2 the dominant uncertainty on $\mathcal{F}_K^{[1.1,6]}$ and $\mathcal{F}_{K^*}^{[1.1,6]}$ stems from the inclusive $B \rightarrow X_s \ell \ell$ measurement which has been recently presented by Belle II [50] using 400 fb^{-1} of integrated luminosity. This measurement is currently statistically limited and, with about 1 ab^{-1} (which is expected by the end of the year) statistical and systematic components of the uncertainty will be about identical. The latter are dominated by the current determinations of the $B \rightarrow K^{(*)} \ell \ell$ branching ratios and by modeling of the hadronic X_s system [50]. A conservative estimate based on 1 ab^{-1} integrated luminosity yields an 8.5% projected uncertainty on the world average. This corresponds to about 10% expected uncertainties on both $\mathcal{F}_K^{[1.1,6]}$ and $\mathcal{F}_{K^*}^{[1.1,6]}$ in the very near future.

At high- q^2 the situation is much brighter because the experimental uncertainties can be improved with LHCb

TABLE III. Theoretical determinations of the ratios $\mathcal{F}_{K^{(*)}}^{[1.1,6],>15}$ for various values of $\mathcal{C}_9^{\text{NP}}$ and $\mathcal{C}_{9,\text{extra}}$ and comparison to their current experimental determinations. Notice that, by construction, the sensitivity to NP is almost flat.

	$\mathcal{F}_K^{>15}$	$\mathcal{F}_{K^*}^{>15}$	$\mathcal{F}_K^{[1.1,6]}$	$\mathcal{F}_{K^*}^{[1.1,6]}$
$(\mathcal{C}_9^{\text{NP}}, \mathcal{C}_{9,\text{extra}}) = (0, 0)$	0.270(64)	0.64(14)	0.099(11)	0.111(60)
$(\mathcal{C}_9^{\text{NP}}, \mathcal{C}_{9,\text{extra}}) = (-1.16, 0)$	0.271(64)	0.65(14)	0.0928(98)	0.113(62)
$(\mathcal{C}_9^{\text{NP}}, \mathcal{C}_{9,\text{extra}}) = (0, -1.16)$	0.210(50)	0.50(11)	0.0764(81)	0.093(51)
experiment	0.309(15)	0.629(21)	0.084(13)	0.110(17)

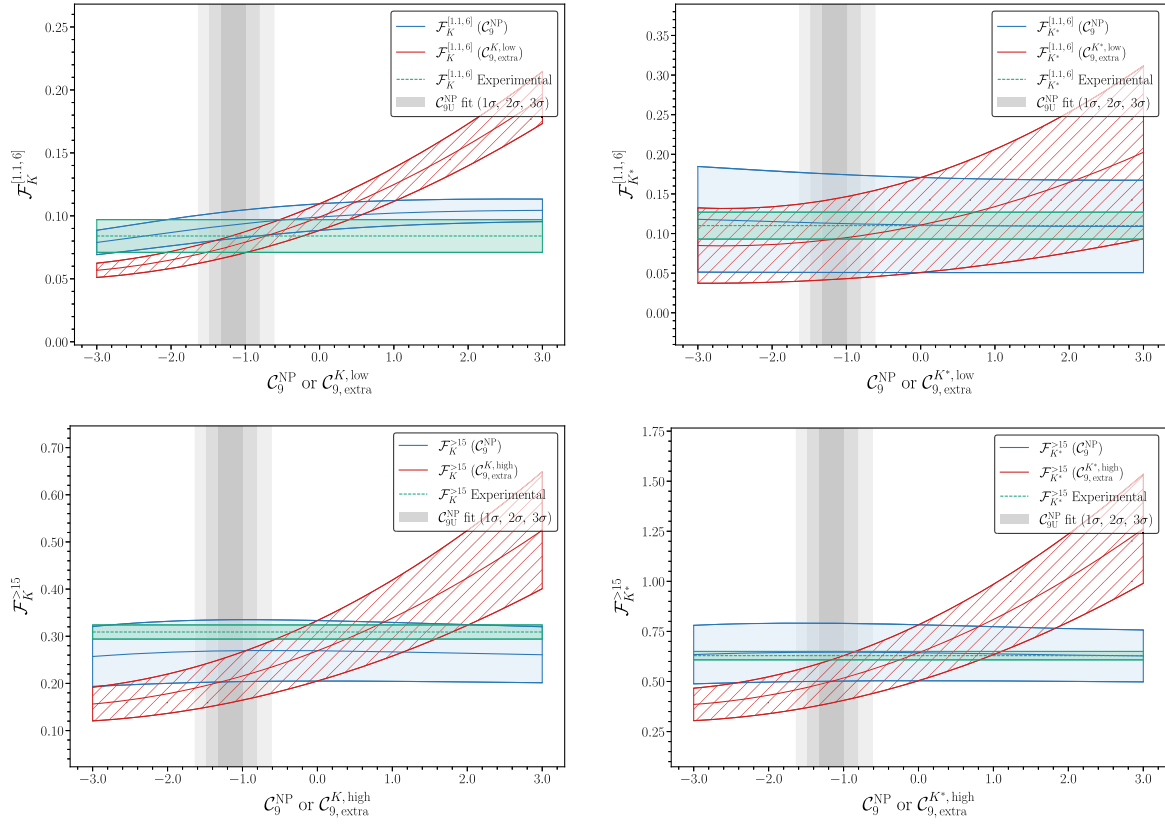


FIG. 2. Current status of experimental and theoretical determinations of the observables $\mathcal{F}_{K^{(*)}}^{[1,1,6],>15}$. The horizontal dashed green lines correspond to the current experimental determinations (and the corresponding 68% CL allowed range) given in Eqs. (16), (17), (18), and (19). The vertical gray bands indicate the 1σ , 2σ , and 3σ results of the global fits to exclusive data at low- and high- q^2 that can be interpreted in terms of C_9^{NP} or $C_{9,\text{extra}}^{K^{(*)}, \text{low or high}}$. The red solid lines (and the corresponding 68% CL allowed range) are the theoretical predictions $\mathcal{F}_{K^{(*)}}^{[1,1,6],>15}(C_9^{\text{NP}} = 0, C_{9,\text{extra}}^{K^{(*)}, \text{low or high}})$. The blue solid line (and the corresponding 68% CL allowed range) is the theoretical prediction $\mathcal{F}_{K^{(*)}}^{[1,1,6],>15}(C_9^{\text{NP}}, C_{9,\text{extra}}^{K^{(*)}, \text{low or high}} = 0)$.

data only. In a first scenario, the $\mathcal{F}$ ratios are measured following the procedure described in Sec. IV, i.e., with improved determinations of the various *normalized* exclusive rates but without any additional measurement of the normalizing charmonium modes. A projection based on a scenario in which the uncertainties on the normalizing charmonium modes are kept at the current level and dominate the exclusive high- q^2 branching ratios that enter the $\mathcal{F}$ ratios, yields a reduction of the overall uncertainties from 4.9% to 3.1% and from 3.3% to 1.4% for $\mathcal{F}_K^{>15}$ and $\mathcal{F}_{K^*}^{>15}$, respectively.

In a second scenario, the $\mathcal{F}$ ratios are constructed directly by LHCb by combining ratios of the various $b \rightarrow s\mu\mu$ exclusive modes. A complication arises because the poor knowledge of the b -quark fragmentation fractions allows only precise measurements of same charge modes, e.g., $\Gamma(B^+ \rightarrow K^{*+}\mu\mu)/\Gamma(B^+ \rightarrow K^+\mu\mu)$ and $\Gamma(B^0 \rightarrow K^{*0}\mu\mu)/\Gamma(B^0 \rightarrow K^0\mu\mu)$. There is, therefore, a trade-off between the reduced uncertainty due to the missing external normalizations and the need to consider modes with neutral

pseudoscalar mesons in the final state. Focusing, for instance, on the K^*/K isospin averaged ratio the two scenarios correspond to

$$\begin{aligned} & \frac{\mathcal{B}(B \rightarrow K^*\mu\mu)}{\mathcal{B}(B \rightarrow K\mu\mu)} \\ &= \begin{cases} \frac{\langle \mathcal{B}(K^{*0}\mu\mu), \mathcal{B}(K^{*+}\mu\mu) \rangle}{\langle \mathcal{B}(K^0\mu\mu), \mathcal{B}(K^+\mu\mu) \rangle} \simeq \frac{\mathcal{B}(K^{*0}\mu\mu)}{\mathcal{B}(K^+\mu\mu)} = \frac{\mathcal{R}(K^{*0}) \mathcal{B}(K^{*0}J/\psi)}{\mathcal{R}(K^+) \mathcal{B}(K^+J/\psi)}, \\ \left\langle \frac{\mathcal{B}(K^{*0}\mu\mu)}{\mathcal{B}(K^0\mu\mu)}, \frac{\mathcal{B}(K^{*+}\mu\mu)}{\mathcal{B}(K^+\mu\mu)} \right\rangle, \end{cases} \quad (20) \end{aligned}$$

where the two ratios in the second scenario do not need any external normalization. In the experimental determination of these two ratios many sources of systematics are expected to cancel; thus an estimate of the projected experimental uncertainties requires a dedicated analysis.

Let us now comment on future prospects for the uncertainties on the theoretical determinations of the $\mathcal{F}$ ratios.

The numerators of these ratios are given by the exclusive $B \rightarrow K^{(*)}\mu\mu$ branching ratios: the only major improvement expected in the near future will come from the calculation of the $B \rightarrow K^*$ form factors in lattice-QCD calculations that take into account the resonant nature of the K^* .⁴

The denominators of these ratios are given by the inclusive $B \rightarrow X_s\ell\ell$ branching ratios at low- and high- q^2 . While the theoretical uncertainty on the branching ratio at low- q^2 is not expected to improve, the high- q^2 result will benefit from three near future developments. As discussed in Secs. II B and III, we construct the high- q^2 branching ratio as the product $\mathcal{R}(B \rightarrow X_s\ell\ell)_{>15} \times \mathcal{B}(B \rightarrow X_u\ell\nu)_{>15}^{\text{exp}}$. First, Belle II is expected to present an improved measurement of the $B \rightarrow X_u\ell\nu$ branching ratio for $q^2 > 15 \text{ GeV}^2$; in a recent analysis [54], Belle II measured the integrated branching ratio for $q^2 > 8 \text{ GeV}^2$ with an uncertainty that is roughly a factor of two better than the corresponding Belle measurement we use. Second, the largest parametric uncertainty in the calculation of the ratio $\mathcal{R}$ is the Cabibbo–Kobayashi–Maskawa (CKM) matrix element $|V_{ub}|$, whose extraction from exclusive and inclusive $b \rightarrow u\ell\nu$ transitions is also expected to improve in the near future, thanks to new lattice calculations of the $B \rightarrow \pi$ form factors and updates to the inclusive determinations at next-to-next-to-leading order accuracy [55,56]. Finally, the single largest uncertainty on $\mathcal{R}$ stems from our poor knowledge of certain vacuum annihilation matrix elements which have been recently calculated, in the D_s sector, using lattice-QCD [57,58]; the corresponding matrix elements for B mesons are more complicated but possible to calculate within the same framework.

In conclusion, it is very likely that the uncertainties on the theoretical determinations of the high- q^2 $\mathcal{F}$ ratios will improve by more than 50% in the near future.

VI. COMPARISON WITH DATA OF INCLUSIVE AND EXCLUSIVE AT HIGH- q^2

At this point, we should remind the reader that one of the main questions we want to address in this paper is: can all observed deviations be explained solely by an unknown universal hadronic effect that mimics NP, or is genuine NP required and unavoidable? Obviously, even if a small non-zero NP contribution is needed to explain all data or part of it, it would imply that a hadronic explanation alone is insufficient, and NP would necessarily be present. Therefore the answer to the question will be in that case no and the presence

of NP will be unavoidable. These observations together with the preference of global fits mentioned in the previous section establish our main starting point: we assume here the presence of a constant universal hadronic contribution $\mathcal{C}_{9,\text{extra}}$ that mimics NP as an explanation of all the anomalies and explore its consequences.

Our complementary strategy here to disentangle a hypothetical hadronic contribution from a potential NP effect is to compare the theoretical prediction for the inclusive rate with the sum of the exclusive decay channels, and confront both with experimental data. This approach should be viewed as an additional, somewhat more rudimentary test than the $\mathcal{F}$ -observables introduced in the previous sections, less precise, perhaps, but still able to provide interesting indications.

In this section, we restrict our attention to the high- q^2 region (above 15 GeV^2), which has the key property that the sum of the branching ratios $\mathcal{B}(B \rightarrow K^*\ell\ell)$ and $\mathcal{B}(B \rightarrow K\ell\ell)$ essentially saturates the inclusive SM prediction. In addition, there is a small residual contribution from the remaining channels, estimated in Ref. [24]. Using the same inputs (Wilson coefficients, quark masses, CKM parameters, normalizations, etc.) employed for the other exclusive channels, we obtain for the dependence of this residual branching ratio on $\mathcal{C}_9^{\text{NP}}$ and on the universal $\mathcal{C}_{9,\text{extra}}$

$$10^7 \times \mathcal{B}_{\text{rest}>15} = 0.547 + 0.124(\mathcal{C}_9^{\text{NP}} + \mathcal{C}_{9,\text{extra}}) + 0.015(\mathcal{C}_9^{\text{NP}} + \mathcal{C}_{9,\text{extra}})^2. \quad (21)$$

This contribution that amounts to 13% of the total sum of exclusive channels is more uncertain than the other exclusive channels, but we find very remarkable and nontrivial that it provides the precise amount needed, as discussed below, to reproduce almost exactly the $\mathcal{C}_9$ -dependence of the inclusive mode. One can also easily see, by inspection of the relative weights of the coefficients (linear versus quadratic) in Eq. (21), that they are similar to the relative weights appearing in the other exclusive channels shown in Tables I and II.

The first step is the comparison between the theoretical expression for the inclusive branching ratio, given in Eq. (12) as a function of $\mathcal{C}_9^{\text{NP}}$ and evaluated following Sec. II B, and the prediction obtained by summing the exclusive decays, evaluated according to Sec. II A. For the latter, we find

$$\begin{aligned} 10^7 \times \mathcal{B}_{\text{exc}} &= 10^7 \times \sum_{\text{mode}} \mathcal{B}_{\text{mode}>15} \\ &= 10^7 \times (\mathcal{B}_{B \rightarrow K\ell\ell>15} + \mathcal{B}_{B \rightarrow K^*\ell\ell>15} + \mathcal{B}_{\text{rest}>15}) \\ &\simeq 4.29 + 0.98(\mathcal{C}_9^{\text{NP}} + \mathcal{C}_{9,\text{extra}}) \\ &\quad + 0.13(\mathcal{C}_9^{\text{NP}} + \mathcal{C}_{9,\text{extra}})^2, \end{aligned} \quad (22)$$

with an uncertainty of $\delta = \pm 0.39$. These results are neatly summarized in Table IV where we show the comparison, as a

⁴A first calculation of the form factors into a light vector meson ($B \rightarrow \rho$) has been presented in Ref. [51] (albeit for a heavy pion, $m_\pi = 320 \text{ MeV}$). The resonance parameters of the K^* have been calculated at the physical point in Refs. [52,53]. Currently, two groups (Jevsenak, Meinel, Leskovec; Black, Boyle, Di Carlo, Erben, Gülpers, Hansen, Lachini, Mukherjee, Portelli, Tsang) are working toward the calculation of the $B \rightarrow K^*$ form factors at the physical point.

TABLE IV. Theoretical predictions for the inclusive $B \rightarrow X_s\mu^+\mu^-$ branching ratios in units of 10^{-7} , obtained summing over the K , K^* and $K\pi$ modes (sum excl) and within an OPE framework (incl). Each branching ratio is written as a quadratic form $\alpha + \beta\tilde{C}_9 + \gamma\tilde{C}_9^2$ where $\tilde{C}_9 = C_9^{\text{NP}} + C_{9,\text{extra}}$ (assuming universality) for the sum of exclusive and $\tilde{C}_9 = C_9^{\text{NP}}$ for the inclusive. In the first three rows we show the breakdown of the sum of exclusive in its three components.

	Region (q^2)	α	β	γ
$10^7 \times \mathcal{B}(B \rightarrow K\mu\mu)$	>15	1.105	0.252	0.033
$10^7 \times \mathcal{B}(B \rightarrow K^*\mu\mu)$	>15	2.640	0.600	0.082
$10^7 \times \mathcal{B}(B \rightarrow K\pi\mu\mu)$	>15	0.547	0.124	0.015
$10^7 \times \mathcal{B}(B \rightarrow X_s\mu\mu)$ (sum excl)	>15	4.292	0.976	0.130
$10^7 \times \mathcal{B}(B \rightarrow X_s\mu\mu)$ (incl)	>15	4.10	0.96	0.14

function of C_9 , between the inclusive branching ratio $B \rightarrow X_s\mu^+\mu^-$ at high- q^2 calculated as a sum of exclusive modes (K , K^* , and $K\pi$) and using the OPE approach. Note the remarkable agreement between the two calculations and how the three exclusive contributions combine, for each power of C_9 , to reproduce the result of the OPE calculation.

The comparison between the inclusive and the sum of exclusive will be done at two different levels, comparing, first, directly the branching ratios and second, comparing the C_9 -dependence of the inclusive versus the exclusive that has a smaller sensitivity on input parameters.

Let us start with the branching ratios. The main point is that the same universal parameter $C_{9,\text{extra}}$ should fulfill two different conditions:

- (i) On the one side, $C_{9,\text{extra}}$ has to be in agreement with experimental data

$$\left(\mathcal{B}_{\text{SM}}^{\text{exc}} - \mathcal{B}_{\text{SM}}^{\text{exc,exp}}\right) = 0, \quad (23)$$

where in the absence of $C_{9,\text{extra}}$ and C_9^{NP} the sum of exclusive exhibits a tension of 3.7σ between data $\mathcal{B}_{\text{SM}}^{\text{exc,exp}}$ given by Eq. (15) and SM. Therefore this would require in the absence of NP that $C_{9,\text{extra}} = -1.16 \pm 0.17$. For the inclusive mode, the deviation from the data amounts to 1.6σ , owing to its larger uncertainty.

- (ii) On the other side, $C_{9,\text{extra}}$ has to be adjusted to provide agreement between the inclusive and the sum of exclusive computation at all points (with or without NP) but particularly, for the SM prediction ($C_9^{\text{NP}} = 0$)

$$\left(\mathcal{B}_{\text{SM}}^{\text{exc}} - \mathcal{B}_{\text{SM}}^{\text{inc}}\right) = 0. \quad (24)$$

Notice that Eqs. (12) and (22) rely on very different computational frameworks, yet their predictions

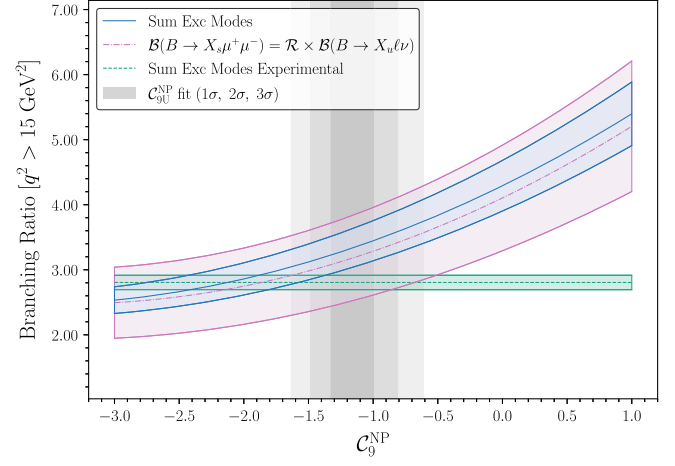


FIG. 3. $B \rightarrow X_s\mu\mu$ branching ratio for $q^2 > 15 \text{ GeV}^2$ calculated using the OPE ($\mathcal{R} \times \mathcal{B}_{u\ell\nu}$) or as a sum of exclusive and comparison with the experimental determination.

agree remarkably well. Figure 3 shows excellent agreement between the inclusive prediction and the sum of exclusive modes in the presence of NP. If we impose that both computations must agree one finds at 68% CL

$$C_{9,\text{extra}} = -0.20_{-1.10}^{+0.90}. \quad (25)$$

This theoretical comparison shows a preference for a small absolute value, compatible with zero, compared with the global fit preferred value of $C_9^{\text{NP}} \sim -1.16 \pm 0.17$ although the uncertainty is still large. Moreover, the profile is highly nonparabolic and the likelihood very asymmetric, i.e., going beyond the 68% CL the range spans up to very large values. Also there is a second solution which again implies a very large negative value (~ -7) that would be in stark contradiction with the preferred values of $C_9^{\text{NP}} \sim -1.16 \pm 0.17$ to explain all the anomalies. Interestingly, the uncertainty in Eq. (25) is mostly driven by the uncertainty in the inclusive computation. If we assume a reduction of 50% in this error as discussed in the previous section and the central value remains untouched one finds at 68% CL

$$C_{9,\text{extra}} = -0.20_{-0.62}^{+0.59}. \quad (26)$$

In this case, the tails of the profile likelihood are more strongly constrained, and at the 95% CL one obtains $C_{9,\text{extra}} = -0.20_{-1.30}^{+1.18}$. Therefore it is very important to improve on the inclusive uncertainty by improving the measurement of $\mathcal{B}(B \rightarrow X_u\ell\nu)$.

To summarize the comparison at the level of branching ratios: the universality pointed by data disfavors the residual extra $C_{9,\text{extra}}$ contribution as a solution of all B -flavor anomalies, once we impose the condition that

the sum of exclusive branching ratios should be equal to the inclusive branching ratio. Therefore, one would need to break this universality, contrary to what the data seem to indicate.

We now turn to the comparison of the C_9 dependencies of the inclusive mode and the sum of exclusive modes. Part of the remaining difference between the inclusive and the sum of exclusive modes arises from small differences in the input parameters used in the two evaluations. To perform a different comparison at the level of C_9 -dependencies in order to reduce normalization-dependent effects, it is convenient to define the following ratios:

$$R^{\text{inc}} = \frac{\mathcal{B}^{\text{inc}}}{\mathcal{B}_{\text{SM}}^{\text{inc}}} = 1 + 0.234C_9^{\text{NP}} + 0.034C_9^{\text{NP}2},$$

$$R^{\sum \text{exc}} = \frac{\mathcal{B}_{\text{SM}}^{\sum \text{exc}}}{\mathcal{B}_{\text{SM}}^{\sum \text{exc}}} = \frac{1 + 0.227(C_9^{\text{NP}} + C_{9,\text{extra}}) + 0.030(C_9^{\text{NP}} + C_{9,\text{extra}})^2}{1 + 0.227C_{9,\text{extra}} + 0.030C_{9,\text{extra}}^2}.$$
(27)

The two limiting scenarios of interest are: (i) $C_9^{\text{NP}} = 0$, the hypothetical hadronic contribution entirely mimics NP, implying $R^{\sum \text{exc}} = 1$, and (ii) $C_{9,\text{extra}} = 0$, the purely NP case. Here we can compare the two expressions in Eq. (27), for which the coefficient of the dominant linear term differs by only 2.8% between the inclusive and exclusive determinations.

The corresponding experimental value is

$$R^{\text{exp}} = \frac{\mathcal{B}_{\text{SM}}^{\sum \text{exc,exp}}}{\mathcal{B}_{\text{SM}}^{\text{inc}}} = 0.68 \pm 0.14, \quad (28)$$

where the uncertainty includes $\delta\mathcal{B}_{\text{SM}}^{\text{inc}}$ from Eq. (12).

This is illustrated in Fig. 4. The two curves corresponding to R^{inc} and $R^{\sum \text{exc}}$ in the NP-only scenario overlap almost perfectly, confirming the consistency between the predictions for the inclusive decay and sum of exclusive channels. Both curves are in perfect agreement with the NP global fit solution (gray bands) of Fig. 4. Instead, the purely hadronic contribution, i.e., $R^{\sum \text{exc}} = 1$ is in tension with the experimental band at 2.3σ . One also can observe in Fig. 4 that when $C_{9,\text{extra}}$ is switched on the tension between the R^{inc} and $R^{\sum \text{exc}}$ increases for large negative values of NP if $C_{9,\text{extra}} \sim -1$ (yellow band). The tiny uncertainty band of the R ratios is a direct consequence of the strong correlation between the theory predictions for the numerator and denominator.

In summary, this alternative approach shows: (i) on the one side, again a preference for NP in front of a constant hadronic universal contribution to explain the data in

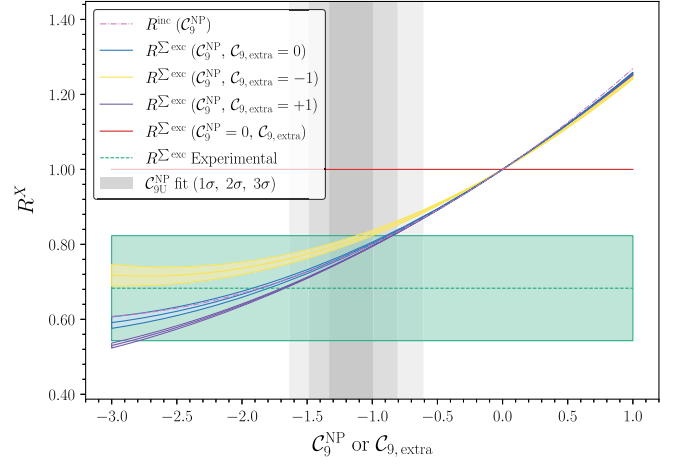


FIG. 4. A more detailed comparison of the C_9 -dependence of the inclusive and exclusive modes under different hypotheses. Green corresponds to the experimental result, R^{exp} , the red line corresponds to the absence of NP with only hadronic contributions, i.e., $R^{\sum \text{exc}}(C_9^{\text{NP}} = 0, C_{9,\text{extra}})$, and the blue, yellow and magenta curves correspond to $R^{\sum \text{exc}}(C_9^{\text{NP}}, C_{9,\text{extra}} = 0)$ and $R^{\sum \text{exc}}(C_9^{\text{NP}}, C_{9,\text{extra}} = \mp 1)$ respectively. This shows that when $C_{9,\text{extra}}$ is switched on (magenta and yellow bands) the necessary good agreement between inclusive and sum of exclusive breaks down.

agreement with what the first analysis of the $\mathcal{F}$ observables (still with large uncertainties) seems to point to, and (ii) on the other side, that a universal unknown hadronic contribution has a different signature from a pure NP contribution, when the inclusive mode is included.

VII. CONCLUSIONS AND OUTLOOK

Since the observation of the first B -flavor anomalies it has become customary to question whether these deviations arise from genuine physics beyond the SM or from unaccounted hadronic effects that reflect our limited understanding of QCD. On the experimental side, the anomalies found in the $b \rightarrow s\mu\mu$ mediated decays have been experimentally confirmed in a persistent and coherent manner, not so in the case of lepton flavor universality violating observables. Moreover, global fits to a large set of semileptonic observables exhibit a clear preference for a single universal NP contribution to the Wilson coefficient of the operator O_9 .

In this analysis our aim has been to discern whether such a NP contribution is unavoidable to explain all the anomalies. Consequently, if a hadronic contribution is able to mimic a universal NP shift in C_9 , it would have to do so consistently across all kinematic regions: in every q^2 bin, both at low and high recoil, and for all decay channels sensitive to this coefficient. This requirement leads to an immediate implication: the hypothetical hadronic contribution must be approximately constant and universal in

order to reproduce the observed NP-like pattern. The proposal that loops involving $D_{(s)}$ and $D^{(*)}$ resonances could account for the observed deviations, if not already included by the corresponding partonic computation, is therefore automatically disfavored. As shown in Refs. [16,17], monopole (dipole) photon couplings generate low- and high- q^2 contributions with opposite (the same) sign. Therefore, large effects at low- q^2 , which require constructive monopole/dipole interference, correlate with much smaller effects at high- q^2 , thus breaking the required universality. A constant and universal hadronic contribution remains a viable possibility.

At this point, a detailed examination of the theoretical computation of the inclusive $B \rightarrow X_s \ell\ell$ decay in contrast with that of the corresponding exclusive channels becomes essential. The crucial feature which we rely on is that, in the inclusive OPE, the effects of charm loops are under control and do not allow for the appearance of nonperturbative effects, like those described by the effective $C_{9,\text{extra}}$ coefficients we introduced above. Factorizable charm loops effects are *exactly* taken into account using $e^+e^- \rightarrow$ hadrons data and the so-called Krüger-Sehgal mechanism [37]. Nonfactorizable effects at low- q^2 are described, within SCET-I, in terms of B -meson subleading shape functions and are estimated conservatively at the 5% level. The situation at high- q^2 is less clear cut. On one side, these nonfactorizable charm effects are given in terms of *local* $O(1/m_c^2)$ power corrections; on the other, the fact that the $q^2 > 15 \text{ GeV}^2$ region integrates over several broad charmonium resonances leaves open the possibility that some additional corrections might be present. Nonetheless, it is important to stress that, while these charm loop effects can, in principle, be modeled for inclusive decays, this is not the case for exclusive modes which rely on a large q^2 expansion.

In this work, motivated by the different sensitivities of inclusive and the corresponding exclusive decays to hadronic contributions, we introduce a new observable, denoted by $\mathcal{F}$, that will, in the near future, allow one to discriminate between the two possibilities: genuine NP or a constant and universal hadronic contribution. Current data exhibit a certain preference for the NP interpretation, although the associated uncertainties are still substantial. There is a large space for improvement especially on the inclusive mode and our projections show that in the high- q^2 region the $\mathcal{F}$ ratios can improve by more than 50%.

An important point that we have stressed is that all current PDG [59] and HFLAV [60] averages for the various modes needed by LHCb and CMS to normalize their branching ratio measurements are currently calculated without taking properly into account the $\Upsilon \rightarrow (B^+B^-, B^0\bar{B}^0)$ branching ratios. This effects have been discussed in the literature [60–62] but they have not been implemented in the averages yet. Since the uncertainties on the various normalizing modes are already a large (and, in

some cases, the largest) fraction of the total uncertainties in LHCb and CMS branching ratios measurements, this issue has to be urgently resolved. In this work, we attempt to correct this problem using a two-pronged approach. First, using the fractions $f^{+,-,00}$ determined in Appendix A, we recalculate the averages for all required charmonium branching ratios. Second, we remove the charmonium normalizations from the LHCb and CMS exclusive $b \rightarrow s\mu^+\mu^-$ branching ratio measurements and reintroduce them using our updated averages.

Finally, at a more basic level, we compared the theoretical computation of the inclusive branching ratio with the one of the sum of exclusive branching ratios. By imposing the equivalence of the two computations, we found that a universal hadronic contribution is compatible with zero and remains rather small (in absolute value) compared to the preferred value provided by the global fits. Although current data still carry uncertainties that are too large to draw definite conclusions, the expected improvement in the precision of the inclusive mode could soon provide a more stringent constraint on this hypothetical universal hadronic contribution. We also showed that one can define a set of SM-normalized observables (R^X with $X = \text{inc}, \sum \text{exc}$) that enable a comparison of the C_9 dependence of the inclusive and sum of exclusive modes with reduced sensitivity to input-parameter choices. This demonstrates, on the one hand, the excellent agreement between the C_9 dependencies of the two computations and, on the other hand, again a preference for a NP interpretation over a universal hadronic contribution.

In summary we have shown that the imprint of a NP or a universal hadronic contribution in the observables once the inclusive mode is considered can be distinguished with a current preference for the NP solution.

ACKNOWLEDGMENTS

The authors are grateful to Kostas Petridis for helpful discussions. J. M. gratefully acknowledges the financial support from ICREA under the ICREA Academia program 2018 and to AGAUR under the Icrea Academia program 2024 and from Departament de Recerca i Universitats de la Generalitat de Catalunya (No. SGR-00649). J. M. also received financial support from the Spanish Ministry of Science, Innovation and Universities (Project No. PID2023-146142NB-I00) and from the IPPP Diva Award 2024. E. L. acknowledges support from CERN where part of this work was performed. The work of B. C. is supported by the La Caixa Junior Leader fellowship from the “la Caixa” Foundation (ID 100010434, fellowship code LCF/BQ/PI24/12040024).

DATA AVAILABILITY

There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

APPENDIX A: BRANCHING RATIOS OF NORMALIZATION MODES

In this appendix we calculate the experimental averages of the branching ratios for $B^{(*)+} \rightarrow K^{(*)+}J/\psi$, $B^{(*)0} \rightarrow K^{(*)0}J/\psi$ and $B^+ \rightarrow K^+\psi(2S)$. Our starting point are the results collected in the PDG [59] or HFLAV [60].

A problem that affects almost all $B^{0,\pm}$ branching ratios measured at B-factories is the impact of $\Upsilon(4S)$ branching ratios to B^+B^- (f^{+-}), $B^0\bar{B}^0$ (f^{00}) and modes without b -hadrons (f_B), with $f^{+-} + f^{00} + f_B = 1$. Currently almost all measured branching fractions which enter the PDG and HFLAV averages have been obtained with the $f^{+-} = f^{00} = 1/2$ assumption.

This problem has been discussed recently in Refs. [61,62]. Here we follow the analysis presented in Sec. 4.1 of the HFLAV review [60] (which includes improvements suggested in Ref. [61]), and combine all existing information on f^{+-}/f^{00} , f^{00} and f_B with a profile log-likelihood fit.

The only departure from the analysis presented in Ref. [60] concerns the calculation of the f^{+-}/f^{00} average which is then fed to the profile log-likelihood fit. First of all, we harmonize the values of the ratio of charged and neutral B mesons used in the various measurements listed in Table 4 of Ref. [60]. This has been done by dividing each measurement by the chosen ratio of lifetimes and multiplying the average by the latest PDG value $\tau(B^+)/\tau(B^0) = 1.075 \pm 0.004$ [59]. A second improvement consists in the estimate of isospin breaking effects. The only f^{+-}/f^{00} measurement which incorporates this estimate is the most recent Belle one [63] where a 4.4% isospin breaking uncertainty is added to the final result, following the analysis of isospin breaking corrections presented in Ref. [64]. This uncertainty is 100% correlated across all f^{+-}/f^{00} measurements; therefore, we remove it from the Belle [63] result, average all measurements and reintroduce the uncertainty on the average. The result of this procedure is $(f^{+-}/f^{00})_{\text{average}} = 1.063 \pm 0.050$.

The f^{+-}/f^{00} average has an uncertainty dominated by isospin breaking ($\pm 4.4\%$ from Ref. [64]) and is considerably less precise than the result presented in Ref. [62] which reads 1.062 ± 0.019 . We prefer to follow the procedure described above as it is a straightforward improvement on the HFLAV analysis and does not modify the estimate of the uncertainty associated with isospin breaking corrections presented in Refs. [63,64].

The profile log-likelihood fit is then identical to that described in Ref. [60]. The two dimensional $\chi^2_{\min} + 1$ and $\chi^2_{\min} + 4$ contours that we obtain are presented in Fig. 5. After profiling over f_B we obtain the following symmetrized averages:

$$f^{+-} = 0.511 \pm 0.010, \quad (\text{A1})$$

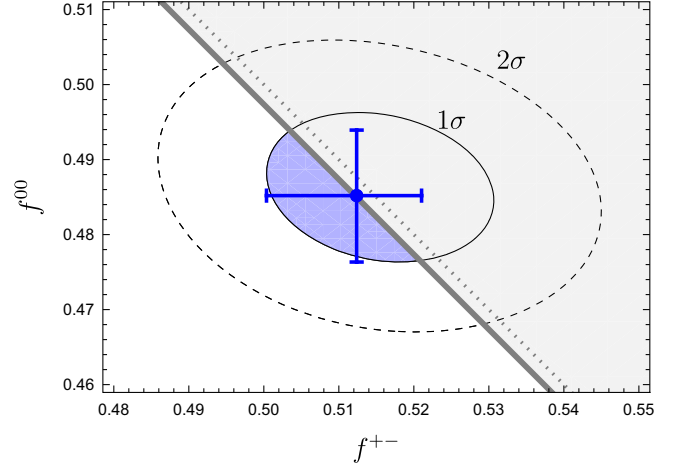


FIG. 5. Result of the profile log-likelihood fit following the strategy presented in Ref. [60]. The two contours correspond to $\chi^2_{\min} + 1$ and $\chi^2_{\min} + 4$. The gray region is excluded by a lower limit on f_B and causes the asymmetry in the fit result (in blue).

$$f^{00} = 0.4851 \pm 0.0088. \quad (\text{A2})$$

For the $B \rightarrow KJ/\psi$ modes, the PDG collects a series of measurements, all of which with the exception of the most recent Belle result assume $f^{+-} = f^{00} = 1/2$. We have verified that the PDG averages are simple weighted averages of these inputs. In order to properly include the effects of the measured $\Upsilon(4S)$ b -hadron fractions, we (1) remove them from the Belle results (using the f^{+-} and f^{00} adopted by Belle), (2) combine the latter with all other measurements, and (3) reintroduce the b -hadron fractions given in Eqs. (A1) and (A2) into the final averages.

In the first step we remove the effect of $f^{+- (00)} \neq 1/2$ by multiplying by $2f^{+- (00)}$ and reducing the uncertainty in quadrature⁵

$$\begin{aligned} \mathcal{B}(B^+ \rightarrow K^+J/\psi)_{\text{Belle}} &= \begin{cases} (10.32 \pm 0.25) \times 10^{-4} & [f^{+-} = 0.514(6)] \\ (10.61 \pm 0.23) \times 10^{-4} & [f^{+-} = 1/2] \end{cases}, \end{aligned} \quad (\text{A3})$$

$$\begin{aligned} \mathcal{B}(B^0 \rightarrow K^0J/\psi)_{\text{Belle}} &= \begin{cases} (9.02 \pm 0.28) \times 10^{-4} & [f^{00} = 0.486(6)] \\ (8.77 \pm 0.25) \times 10^{-4} & [f^{00} = 1/2] \end{cases}. \end{aligned} \quad (\text{A4})$$

Then we average these branching ratios with all those included in the PDG average and reintroduce the fractions

⁵If the number of $B^+ \rightarrow K^+J/\psi$ events is N^+ and the total number of BB pairs is N the inclusion of the ratio f^{+-} causes the branching ratio to shift from $\mathcal{B}^{50:50} = N^+/(N/2)$ to $\mathcal{B}^{\text{correct}} = N^+/(Nf^{+-}) = \mathcal{B}^{50:50}/(2f^{+-})$.

$f^{+- (00)}$ in the final average. The results are⁶

$$\mathcal{B}(B^+ \rightarrow K^+ J/\psi)_{\text{avg}} = (10.18 \pm 0.28) \times 10^{-4}, \quad (\text{A5})$$

$$\mathcal{B}(B^0 \rightarrow K^0 J/\psi)_{\text{avg}} = (9.03 \pm 0.26) \times 10^{-4}. \quad (\text{A6})$$

The isospin averaged branching ratio we obtain from Eqs. (A5) and (A6) is $9.57(19) \times 10^{-4}$.

For the $K^{*+ (0)} J/\psi$ modes the situation is somewhat simpler because all measurements that enter the PDG average assume $f^{+-} = f^{00} = 1/2$. Therefore we just need to rescale the PDG averages by $1/(2f^{+- (00)})^7$

$$\mathcal{B}(B^+ \rightarrow K^{*+} J/\psi)_{\text{avg}} = (13.97 \pm 0.85) \times 10^{-4}, \quad (\text{A7})$$

$$\mathcal{B}(B^0 \rightarrow K^{*0} J/\psi)_{\text{avg}} = (13.23 \pm 0.54) \times 10^{-4}. \quad (\text{A8})$$

Finally we need the $B^+ \rightarrow K^+ \psi(2S)$ branching ratio. Here the situation is more complex because the PDG combination is not a simple average but is obtained via a more complex fit. Since, all measurements that enter the fit assume $f^{+-} = 1/2$ we simply rescale their final result by $1/(2f^{+-})$:

$$\mathcal{B}(B^+ \rightarrow K^+ \psi(2S))_{\text{avg}} = (6.11 \pm 0.24) \times 10^{-4}. \quad (\text{A9})$$

APPENDIX B: LOW- q^2

In this appendix we calculate the averages for the normalized low- q^2 exclusive $b \rightarrow s\mu\mu$ branching ratios (Sec. B 1) and then combine them with the normalizations calculated in Appendix A (Sec. B 2).

1. $B \rightarrow K^{(*)} \mu\mu$ normalized rates

The four $B \rightarrow K^{(*)} \mu\mu$ decays have been measured for low and high- q^2 by LHCb [65,66]. The charged $B^+ \rightarrow K^+ \mu\mu$ mode has been measured also by CMS [67]. Note that CMS presents results in narrow 1 GeV² bins which we have to combine paying attention not to dilute the uncertainty from the common $B^+ \rightarrow K^+ J/\psi$ normalization. We remove the normalizations adopted by LHCb and CMS and calculate the ratios

$$\mathcal{R}(K^{+(0)}) \equiv \mathcal{B}(B^{+(0)} \rightarrow K^{+(0)} \mu\mu) / \mathcal{B}(B^{+(0)} \rightarrow K^{+(0)} J/\psi), \quad (\text{B1})$$

$$\mathcal{R}(K^{*+ (0)}) \equiv \mathcal{B}(B^{+(0)} \rightarrow K^{*+ (0)} \mu\mu) / \mathcal{B}(B^{+(0)} \rightarrow K^{*+ (0)} J/\psi). \quad (\text{B2})$$

⁶The PDG (HFLAV) averages in units of 10^{-4} for the K^+ and K^0 modes are 10.20 ± 0.19 (10.18 ± 0.18) and 8.91 ± 0.21 (8.83 ± 0.20), respectively.

⁷The PDG (HFLAV) averages in units of 10^{-4} for the K^{*+} and K^{*0} modes are 14.30 ± 0.81 (13.76 ± 0.82) and 12.65 ± 0.46 (12.78 ± 0.44), respectively.

For the K^+ case, we average LHCb and CMS. The results we obtain for the ratios $\mathcal{R}(K^{+(0)})$ are as follows. Note that, in order to remove the normalization, we need to use exactly what the experiments used.

(i) $B^+ \rightarrow K^+ \mu\mu$ (LHCb [65], CMS [67]):

$$\begin{aligned} \mathcal{B}(B^+ \rightarrow K^+ \mu\mu)_{\text{LHCb}} &= (24.2 \pm 0.7 \pm 1.2) \times (6 - 1.1) \times 10^{-9} \\ &= (1.186 \pm 0.068) \times 10^{-7}, \end{aligned} \quad (\text{B3})$$

$$\begin{aligned} \mathcal{B}(B^+ \rightarrow K^+ J/\psi)_{\text{LHCb}} &= (9.98 \pm 0.14 \pm 0.40) \times 10^{-4} \\ &= (9.98 \pm 0.42) \times 10^{-4}, \end{aligned} \quad (\text{B4})$$

$$\mathcal{R}(K^+)_{\text{LHCb}} = (1.188 \pm 0.046) \times 10^{-4}, \quad (\text{B5})$$

$$\mathcal{B}(B^+ \rightarrow K^+ \mu\mu)_{\text{CMS}} = (1.253 \pm 0.057) \times 10^{-7}, \quad (\text{B6})$$

$$\mathcal{B}(B^+ \rightarrow K^+ J/\psi)_{\text{CMS}} = (10.20 \pm 0.19) \times 10^{-4}, \quad (\text{B7})$$

$$\mathcal{R}(K^+)_{\text{CMS}} = (1.228 \pm 0.051) \times 10^{-4}, \quad (\text{B8})$$

$$\mathcal{R}(K^+)_{\text{avg}}^{\text{low}} = (1.206 \pm 0.034) \times 10^{-4}. \quad (\text{B9})$$

(ii) $B^0 \rightarrow K^0 \mu\mu$ (LHCb [65]):

$$\begin{aligned} \mathcal{B}(B^0 \rightarrow K^0 \mu\mu)_{\text{LHCb}} &= (18.85 \pm 3.35 \pm 0.9) \\ &\quad \times (6 - 1.1) \times 10^{-9} \\ &= (0.92 \pm 0.17) \times 10^{-7}, \end{aligned} \quad (\text{B10})$$

$$\begin{aligned} \mathcal{B}(B^0 \rightarrow K^0 J/\psi)_{\text{LHCb}} &= (9.28 \pm 0.13 \pm 0.37) \times 10^{-4} \\ &= (9.28 \pm 0.39) \times 10^{-4}, \end{aligned} \quad (\text{B11})$$

$$\mathcal{R}(K^0)_{\text{avg}}^{\text{low}} = (0.995 \pm 0.18) \times 10^{-4}. \quad (\text{B12})$$

(iii) $B^+ \rightarrow K^{*+} \mu\mu$ (LHCb [65]):

$$\begin{aligned} \mathcal{B}(B^+ \rightarrow K^{*+} \mu\mu) &= (36.95 \pm 7.95 \pm 2.6) \times (6 - 1.1) \times 10^{-9} \\ &= (1.81 \pm 0.41) \times 10^{-7}, \end{aligned} \quad (\text{B13})$$

$$\begin{aligned} \mathcal{B}(B^+ \rightarrow K^{*+} J/\psi) &= (1.431 \pm 0.027 \pm 0.090) \times 10^{-3} \\ &= (1.431 \pm 0.094) \times 10^{-3}, \end{aligned} \quad (\text{B14})$$

$$\mathcal{R}(K^{*+})_{\text{avg}}^{\text{low}} = (1.265 \pm 0.27) \times 10^{-4}. \quad (\text{B15})$$

(iv) $B^0 \rightarrow K^{*0}\mu\mu$ (LHCb [66]):

$$\begin{aligned}\mathcal{B}(B^0 \rightarrow K^{*0}\mu\mu) &= (0.342 \pm 0.017 \pm 0.009 \pm 0.023) \\ &\times (6 - 1.1) \times 10^{-7} \\ &= (1.68 \pm 0.15) \times 10^{-7},\end{aligned}\quad (\text{B16})$$

$$\begin{aligned}\mathcal{B}(B^0 \rightarrow K^{*0}J\psi) &= (1.19 \pm 0.01 \pm 0.08) \times 10^{-3} \\ &= (1.19 \pm 0.081) \times 10^{-3},\end{aligned}\quad (\text{B17})$$

$$\mathcal{R}(K^{*0})_{\text{avg}}^{\text{low}} = (1.408 \pm 0.078) \times 10^{-4}.\quad (\text{B18})$$

Observe that the latest LHCb measurement of the $B^0 \rightarrow K^{*0}\mu\mu$ branching ratio [1] adopts a normalization which is adjusted to the $m_{K^+\pi^-} \in [0.7459, 1.0959]$ GeV mass window. Since this quantity can only be reconstructed from the Belle [68] measurement, we refrain from calculating an updated result for the ratio $\mathcal{R}(K^{*0})$ and present directly the branching ratios given in Ref. [1].

2. $B \rightarrow K^{(*)}\mu\mu$ branching ratios

Combining the results in Appendix A and Sec. B 1 we obtain the following branching ratios:

$$\mathcal{B}(B^+ \rightarrow K^+\mu\mu)_{\text{avg}}^{\text{low}} = (1.228 \pm 0.048) \times 10^{-7},\quad (\text{B19})$$

$$\mathcal{B}(B^0 \rightarrow K^0\mu\mu)_{\text{avg}}^{\text{low}} = (0.90 \pm 0.16) \times 10^{-7},\quad (\text{B20})$$

$$\mathcal{B}(B^+ \rightarrow K^{*+}\mu\mu)_{\text{avg}}^{\text{low}} = (1.77 \pm 0.40) \times 10^{-7},\quad (\text{B21})$$

$$\mathcal{B}(B^0 \rightarrow K^{*0}\mu\mu)_{\text{avg}}^{\text{low}} = (1.578 \pm 0.081) \times 10^{-7},\quad (\text{B22})$$

where the last branching ratio is taken directly from Ref. [1].

The isospin average rates, which are the quantities that appear in the ratios $\mathcal{F}_{K^{(*)}}^{[1.1,6]}$ are

$$\mathcal{B}(B \rightarrow K\mu\mu)_{\text{avg}}^{\text{low}} = (1.202 \pm 0.046) \times 10^{-7},\quad (\text{B23})$$

$$\mathcal{B}(B \rightarrow K^*\mu\mu)_{\text{avg}}^{\text{low}} = (1.585 \pm 0.079) \times 10^{-7}.\quad (\text{B24})$$

Note that the isospin averaged branching ratios do not assume the isospin equality of charged and neutral $B \rightarrow K^{(*)}\mu\mu$ partial widths. Moreover, we have explicitly included an estimate of isospin breaking effects in the f^{+-} and f^{00} determinations.

APPENDIX C: HIGH- q^2

In this appendix we calculate the averages for the normalized high- q^2 exclusive $b \rightarrow s\mu\mu$ branching ratios (Sec. C 1) and then combine them with the normalizations calculated in Appendix A (Sec. C 2).

1. $B \rightarrow K^{(*)}\mu\mu$ normalized rates

We follow the exact same procedure as in Sec. B 1 to calculate the corresponding $\mathcal{R}(K^*)$ for $q^2 > 15$ GeV².

(i) $B^+ \rightarrow K^+\mu\mu$ (LHCb [65], CMS [67]):

$$\begin{aligned}\mathcal{B}(B^+ \rightarrow K^+\mu\mu) &= (12.1 \pm 0.4 \pm 0.6) \\ &\times (22 - 15) \times 10^{-9} \\ &= (0.847 \pm 0.050) \times 10^{-7},\end{aligned}\quad (\text{C1})$$

$$\mathcal{B}(B^+ \rightarrow K^+J\psi) = (9.98 \pm 0.14 \pm 0.40) \times 10^{-4},\quad (\text{C2})$$

$$\mathcal{R}(K^+)_{\text{LHCb}} = (0.849 \pm 0.035) \times 10^{-4},\quad (\text{C3})$$

$$\mathcal{B}(B^+ \rightarrow K^+\mu\mu) = (0.927 \pm 0.046) \times 10^{-7},\quad (\text{C4})$$

$$\mathcal{B}(B^+ \rightarrow K^+J\psi) = (10.20 \pm 0.19) \times 10^{-4},\quad (\text{C5})$$

$$\mathcal{R}(K^+)_{\text{CMS}} = (0.909 \pm 0.042) \times 10^{-4},\quad (\text{C6})$$

$$\mathcal{R}(K^+)_{\text{avg}}^{\text{high}} = (0.874 \pm 0.027) \times 10^{-4}.\quad (\text{C7})$$

(ii) $B^0 \rightarrow K^0\mu\mu$ (LHCb [65]):

$$\begin{aligned}\mathcal{B}(B^0 \rightarrow K^0\mu\mu) &= (9.55 \pm 1.55 \pm 0.5) \times (22 - 15) \times 10^{-9} \\ &= (0.67 \pm 0.11) \times 10^{-7},\end{aligned}\quad (\text{C8})$$

$$\begin{aligned}\mathcal{B}(B^0 \rightarrow K^0J\psi) &= (9.28 \pm 0.13 \pm 0.37) \times 10^{-4} \\ &= (9.28 \pm 0.39) \times 10^{-4},\end{aligned}\quad (\text{C9})$$

$$\mathcal{R}(K^0)_{\text{avg}}^{\text{high}} = (0.72 \pm 0.12) \times 10^{-4}.\quad (\text{C10})$$

(iii) $B^+ \rightarrow K^{*+}\mu\mu$ (LHCb [65]):

$$\begin{aligned}\mathcal{B}(B^+ \rightarrow K^{*+}\mu\mu) &= (39.85 \pm 7.65 \pm 2.8) \times (19 - 15) \times 10^{-9} \\ &= (1.59 \pm 0.33) \times 10^{-7},\end{aligned}\quad (\text{C11})$$

$$\begin{aligned}\mathcal{B}(B^+ \rightarrow K^{*+}J\psi) &= (14.31 \pm 0.27 \pm 0.90) \times 10^{-4} \\ &= (14.31 \pm 0.94) \times 10^{-4},\end{aligned}\quad (\text{C12})$$

$$\mathcal{R}(K^{*+})_{\text{avg}}^{\text{high}} = (1.11 \pm 0.22) \times 10^{-4}.\quad (\text{C13})$$

(iv) $B^0 \rightarrow K^{*0}\mu\mu$ (LHCb [66]):

$$\begin{aligned} \mathcal{B}(B^0 \rightarrow K^{*0}\mu\mu) &= (0.4355 \pm 0.0185 \pm 0.007 \pm 0.030) \\ &\times (19 - 15) \times 10^{-7} \\ &= (1.74 \pm 0.14) \times 10^{-7}, \end{aligned} \quad (\text{C14})$$

$$\begin{aligned} \mathcal{B}(B^0 \rightarrow K^{*0}J\psi) &= (1.19 \pm 0.01 \pm 0.08) \times 10^{-3} \\ &= (1.19 \pm 0.081) \times 10^{-3}, \end{aligned} \quad (\text{C15})$$

$$\mathcal{R}(K^{*0})_{\text{avg}}^{\text{high}} = (1.464 \pm 0.069) \times 10^{-4}. \quad (\text{C16})$$

Similarly to the low- q^2 case, the latest LHCb measurement of the $B^0 \rightarrow K^{*0}\mu\mu$ branching ratio [1] adopts a normalization which is adjusted to the $m_{K^+\pi^-} \in [0.7459, 1.0959]$ GeV mass window. We refrain from calculating an updated result for the ratio $\mathcal{R}(K^{*0})$ and present directly the branching ratios given in Ref. [1].

(v) $B^+ \rightarrow K^+\pi^+\pi^-\mu\mu$ (LHCb [69]):

$$\begin{aligned} \mathcal{B}(B^+ \rightarrow K^+\pi^+\pi^-\mu\mu) &= (19 - 15)(0.11 \pm 0.07 \pm 0.01) \times 10^{-8} \\ &= (0.44 \pm 0.28) \times 10^{-8}, \end{aligned} \quad (\text{C17})$$

$$\mathcal{B}(B^+ \rightarrow K^+\psi(2S)) = (6.27 \pm 0.24) \times 10^{-4}, \quad (\text{C18})$$

$$\mathcal{R}(K^+\pi^+\pi^-) = (0.70 \pm 0.45) \times 10^{-5}. \quad (\text{C19})$$

2. $B \rightarrow K^{(*)}\mu\mu$ branching ratios

Combining the results in Appendix A and Sec. C 1 we obtain the following branching ratios:

$$\mathcal{B}(B^+ \rightarrow K^+\mu\mu)_{\text{avg}}^{\text{high}} = (0.890 \pm 0.036) \times 10^{-7}, \quad (\text{C20})$$

$$\mathcal{B}(B^0 \rightarrow K^0\mu\mu)_{\text{avg}}^{\text{high}} = (0.65 \pm 0.11) \times 10^{-7}, \quad (\text{C21})$$

$$\mathcal{B}(B \rightarrow K\mu\mu)_{\text{avg}}^{\text{high}} = (0.866 \pm 0.035) \times 10^{-7}, \quad (\text{C22})$$

$$\mathcal{B}(B^+ \rightarrow K^{*+}\mu\mu)_{\text{avg}}^{\text{high}} = (1.56 \pm 0.32) \times 10^{-7}, \quad (\text{C23})$$

$$\mathcal{B}(B^0 \rightarrow K^{*0}\mu\mu)_{\text{avg}}^{\text{high}} = (1.780 \pm 0.091) \times 10^{-7}, \quad (\text{C24})$$

$$\mathcal{B}(B \rightarrow K^*\mu\mu)_{\text{avg}}^{\text{high}} = (1.763 \pm 0.088) \times 10^{-7}, \quad (\text{C25})$$

$$\mathcal{B}(B^- \rightarrow (K\pi)_S\mu\mu)_{\text{avg}}^{\text{high}} = (0.11 \pm 0.04) \times 10^{-7}, \quad (\text{C26})$$

$$\mathcal{B}(B^- \rightarrow K^+\pi^+\pi^-\mu\mu)_{\text{avg}}^{\text{high}} = (0.43 \pm 0.28) \times 10^{-8}, \quad (\text{C27})$$

where the K^* and $(K^+\pi^-)_S$ branching ratios are taken directly from Ref. [1]. Notice that the S -wave branching ratio in Eq. (C26) contains a $3/2$ isospin factor required to take into account the $K_S\pi^0$ mode.

Moreover, assuming isospin invariance of the $B^{+,0} \rightarrow (K\pi)_S\mu\mu$ partial widths, we can write the following isospin average:

$$\mathcal{B}(B \rightarrow (K\pi)_S\mu\mu) = \mathcal{B}(B^0 \rightarrow (K\pi)_S\mu\mu) \frac{\tau^0 + \tau^+}{2\tau_0} \quad (\text{C28})$$

$$= (0.114 \pm 0.042) \times 10^{-7}. \quad (\text{C29})$$

Finally, using isospin once more we can obtain an expression for the isospin averaged $B \rightarrow K\pi\pi\mu\mu$ branching ratio

$$\mathcal{B}(B \rightarrow K\pi\pi\mu\mu) = \frac{\mathcal{B}(B^+ \rightarrow K\pi\pi\mu\mu)}{\tau^+} \frac{\tau^0 + \tau^+}{2} \quad (\text{C30})$$

$$= \frac{3\tau^0 + \tau^+}{2} \mathcal{B}(B^+ \rightarrow K^+\pi^+\pi^-\mu\mu) \quad (\text{C31})$$

$$= (0.062 \pm 0.040) \times 10^{-7}, \quad (\text{C32})$$

where we used $\tau^0/\tau^+ = 1.076 \pm 0.004$ [59]. The explicit factor of $3/2$ in Eq. (C31) takes into account modes involving neutral K and π (see the discussion in Sec. 3.2 of Ref. [25]).

-
- [1] LHCb Collaboration, Comprehensive analysis of the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decay, *Phys. Rev. Lett.* **137**, 021802 (2026).
[2] CMS Collaboration, Angular analysis of the $B^0 \rightarrow K^*(892)^0\mu^+\mu^-$ decay in proton-proton collisions at $\sqrt{s} = 13$ TeV, *Phys. Lett. B* **864**, 139406 (2025).
[3] A. Khodjamirian, T. Mannel, A. A. Pivovarov, and Y. M. Wang, Charm-loop effect in $B \rightarrow K^{(*)}\ell^+\ell^-$ and $B \rightarrow K^*\gamma$, *J. High Energy Phys.* **09** (2010) 089.

- [4] N. Gubernari, A. Kokulu, and D. van Dyk, $B \rightarrow P$ and $B \rightarrow V$ form factors from B -meson light-cone sum rules beyond leading twist, *J. High Energy Phys.* **01** (2019) 150.
[5] A. Bharucha, D. M. Straub, and R. Zwicky, $B \rightarrow V\ell^+\ell^-$ in the standard model from light-cone sum rules, *J. High Energy Phys.* **08** (2016) 098.
[6] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, Rare B decays using lattice QCD form factors, *Proc. Sci. LATTICE2014* (2015) 372 [arXiv:1501.00367].

- [7] HPQCD Collaboration, $B \rightarrow K$ and $D \rightarrow K$ form factors from fully relativistic lattice QCD, *Phys. Rev. D* **107**, 014510 (2023).
- [8] M. Ciuchini, M. Fedele, E. Franco, A. Paul, L. Silvestrini, and M. Valli, Constraints on lepton universality violation from rare B decays, *Phys. Rev. D* **107**, 055036 (2023).
- [9] M. Algueró, A. Biswas, B. Capdevila, S. Descotes-Genon, J. Matias, and M. Novoa-Brunet, To (b)e or not to (b)e: No electrons at LHCb, *Eur. Phys. J. C* **83**, 648 (2023).
- [10] T. Hurth, F. Mahmoudi, Y. Monceaux, and S. Neshatpour, Data-driven analyses and model-independent fits for present $b \rightarrow s\ell\ell$ results, *Phys. Rev. D* **112**, 113003 (2025).
- [11] N. Gubernari, M. Reboud, D. van Dyk, and J. Virto, Improved theory predictions and global analysis of exclusive $b \rightarrow s\mu^+\mu^-$ processes, *J. High Energy Phys.* **09** (2022) 133.
- [12] N. Gubernari, M. Reboud, D. van Dyk, and J. Virto, Dispersive analysis of $B \rightarrow K^{(*)}$ and $B_s \rightarrow \phi$ form factors, *J. High Energy Phys.* **12** (2023) 153.
- [13] J. Matias, F. Mescia, M. Ramon, and J. Virto, Complete anatomy of $\bar{B}_d \rightarrow \bar{K}^{*0}(\rightarrow K\pi)l^+l^-$ and its angular distribution, *J. High Energy Phys.* **04** (2012) 104.
- [14] S. Descotes-Genon, T. Hurth, J. Matias, and J. Virto, Optimizing the basis of $B \rightarrow K^*ll$ observables in the full kinematic range, *J. High Energy Phys.* **05** (2013) 137.
- [15] HPQCD Collaboration, Standard model predictions for $B \rightarrow K\ell^+\ell^-$, $B \rightarrow K\ell_1^-\ell_2^+$ and $B \rightarrow K\nu\bar{\nu}$ using form factors from $N_f = 2 + 1 + 1$ lattice QCD, *Phys. Rev. D* **107**, 014511 (2023).
- [16] G. Isidori, Z. Polonsky, and A. Tinari, Explicit estimate of charm rescattering in $B^0 \rightarrow K^0\bar{\ell}\ell$, *Phys. Rev. D* **111**, 093007 (2025).
- [17] G. Isidori, Z. Polonsky, and A. Tinari, Charm rescattering in $B^0 \rightarrow K^0\bar{\ell}\ell$: An improved analysis, *Eur. Phys. J. C* **85**, 1221 (2025).
- [18] H. M. Asatrian, C. Greub, and J. Virto, Exact NLO matching and analyticity in $b \rightarrow s\ell\ell$, *J. High Energy Phys.* **04** (2020) 012.
- [19] M. Beylich, G. Buchalla, and T. Feldmann, Theory of $B \rightarrow K^{(*)}\ell^+\ell^-$ decays at high q^2 : OPE and quark-hadron duality, *Eur. Phys. J. C* **71**, 1635 (2011).
- [20] M. Bordone, G. Isidori, S. Mächler, and A. Tinari, Short- vs long-distance physics in $B \rightarrow K^{(*)}\ell^+\ell^-$: A data-driven analysis, *Eur. Phys. J. C* **84**, 547 (2024).
- [21] M. Hoferichter, B. Kubis, and S. Mutke, Reconciling hadronic and partonic analyticity in $b \rightarrow s\ell\ell$ transitions, *Phys. Lett. B* **881**, 140841 (2026).
- [22] W. Altmannshofer, S. G. Christensen, and P. Stangl, Large Hadronic Effects in $B \rightarrow K^*\mu\mu$?, *arXiv:2603.27753*.
- [23] T. Huber, T. Hurth, J. Jenkins, E. Lunghi, Q. Qin, and K. K. Vos, Phenomenology of inclusive $\bar{B} \rightarrow X_s\ell^+\ell^-$ for the Belle II era, *J. High Energy Phys.* **10** (2020) 088.
- [24] G. Isidori, Z. Polonsky, and A. Tinari, Semi-inclusive $b \rightarrow s\ell^-\ell$ transitions at high q^2 , *Phys. Rev. D* **108**, 093008 (2023).
- [25] T. Huber, T. Hurth, J. Jenkins, E. Lunghi, Q. Qin, and K. K. Vos, Inclusive $\bar{B} \rightarrow X_s\ell^+\ell^-$ at the LHC: Theory predictions and new-physics reach, *J. High Energy Phys.* **11** (2024) 130.
- [26] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, Weak decays beyond leading logarithms, *Rev. Mod. Phys.* **68**, 1125 (1996).
- [27] B. Grinstein, R. P. Springer, and M. B. Wise, Effective Hamiltonian for weak radiative B meson decay, *Phys. Lett. B* **202**, 138 (1988).
- [28] B. Capdevila, A. Crivellin, S. Descotes-Genon, J. Matias, and J. Virto, Patterns of new physics in $b \rightarrow s\ell^+\ell^-$ transitions in the light of recent data, *J. High Energy Phys.* **01** (2018) 093.
- [29] A. Hardmeier, E. Lunghi, D. Pirjol, and D. Wyler, Subleading collinear operators and their matrix elements, *Nucl. Phys. B* **682**, 150 (2004).
- [30] T. Becher, R. J. Hill, and M. Neubert, Soft collinear messengers: A new mode in soft collinear effective theory, *Phys. Rev. D* **69**, 054017 (2004).
- [31] S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto, Global analysis of $b \rightarrow s\ell\ell$ anomalies, *J. High Energy Phys.* **06** (2016) 092.
- [32] M. Algueró, B. Capdevila, A. Crivellin, S. Descotes-Genon, P. Masjuan, J. Matias, and J. Virto, Emerging patterns of new physics with and without lepton flavor universal contributions, *Eur. Phys. J. C* **79**, 714 (2019).
- [33] M. Beneke, T. Feldmann, and D. Seidel, Systematic approach to exclusive $B \rightarrow V l^+ l^-$, $V\gamma$ decays, *Nucl. Phys. B* **612**, 25 (2001).
- [34] C. Bobeth, M. Chrzaszcz, D. van Dyk, and J. Virto, Long-distance effects in $B \rightarrow K^*\ell\ell$ from analyticity, *Eur. Phys. J. C* **78**, 451 (2018).
- [35] N. Gubernari, D. van Dyk, and J. Virto, Nonlocal matrix elements in $B_{(s)} \rightarrow \{K^{(*)}, \phi\}\ell^+\ell^-$, *J. High Energy Phys.* **02** (2021) 088.
- [36] B. Grinstein and D. Pirjol, Exclusive rare $B \rightarrow K^*\ell^+\ell^-$ decays at low recoil: Controlling the long-distance effects, *Phys. Rev. D* **70**, 114005 (2004).
- [37] F. Kruger and L. M. Sehgal, Lepton polarization in the decays $b \rightarrow X_d\mu^+\mu^-$ and $B \rightarrow X_s\tau^+\tau^-$, *Phys. Lett. B* **380**, 199 (1996).
- [38] C. Bobeth, G. Hiller, and D. van Dyk, The benefits of $\bar{B} \rightarrow \bar{K}^*l^+l^-$ decays at low recoil, *J. High Energy Phys.* **07** (2010) 098.
- [39] M. Beneke, G. Buchalla, M. Neubert, and C. T. Sachrajda, Penguins with charm and quark-hadron duality, *Eur. Phys. J. C* **61**, 439 (2009).
- [40] G. Buchalla, G. Isidori, and S. J. Rey, Corrections of order $\Lambda_{\text{QCD}}^2/m_c^2$ to inclusive rare B decays, *Nucl. Phys. B* **511**, 594 (1998).
- [41] T. Hurth, M. Fickinger, S. Turczyk, and M. Benzke, Resolved power corrections to the inclusive decay $\bar{B} \rightarrow X_s\ell^+\ell^-$, *Nucl. Part. Phys. Proc.* **285–286**, 57 (2017).
- [42] M. Benzke, T. Hurth, and S. Turczyk, Subleading power factorization in $\bar{B} \rightarrow X_s\ell^+\ell^-$, *J. High Energy Phys.* **10** (2017) 031.
- [43] M. Benzke and T. Hurth, Resolved $1/m_b$ contributions to $\bar{B} \rightarrow X_{s,d}\ell^+\ell^-$ and $\bar{B} \rightarrow X_s\gamma$, *Phys. Rev. D* **102**, 114024 (2020).
- [44] Z. Ligeti and F. J. Tackmann, Precise predictions for $B \rightarrow X_sl^+l^-$ in the large q^2 region, *Phys. Lett. B* **653**, 404 (2007).

- [45] Belle Collaboration, Measurement of differential branching fractions of inclusive $B \rightarrow X_u \ell^+ \nu_\ell$ decays, *Phys. Rev. Lett.* **127**, 261801 (2021).
- [46] BABAR Collaboration, Measurement of the $B \rightarrow X_s \ell^+ \ell^-$ branching fraction with a sum over exclusive modes, *Phys. Rev. Lett.* **93**, 081802 (2004).
- [47] BABAR Collaboration, Measurement of the $B \rightarrow X_s l^+ l^-$ branching fraction and search for direct CP violation from a sum of exclusive final states, *Phys. Rev. Lett.* **112**, 211802 (2014).
- [48] Belle Collaboration, Improved measurement of the electro-weak penguin process $B \rightarrow X_s l^+ l^-$, *Phys. Rev. D* **72**, 092005 (2005).
- [49] Belle Collaboration, Measurement of the lepton forward-backward asymmetry in $B \rightarrow X_s \ell^+ \ell^-$ decays with a sum of exclusive modes, *Phys. Rev. D* **93**, 032008 (2016).
- [50] Z. Wang (Belle-II Collaboration), Measurement of inclusive $B \rightarrow X_s \ell^+ \ell^-$ branching fractions at Belle II, in *Talk at 2025 Rencontres de Moriond, Electroweak Interactions and Unified Theories*.
- [51] L. Leskovec, S. Meinel, M. Petschlies, J. Negele, S. Paul, and A. Pochinsky, $B \rightarrow \rho \ell \nu^-$ resonance form factors from $B \rightarrow \pi \pi \ell \nu^-$ in lattice QCD, *Phys. Rev. Lett.* **134**, 161901 (2025).
- [52] P. Boyle, F. Erben, V. Gülpers, M. T. Hansen, F. Joswig, M. Marshall, N. P. Lachini, and A. Portelli, Light and strange vector resonances from lattice QCD at physical quark masses, *Phys. Rev. Lett.* **134**, 111901 (2025).
- [53] Q.-Z. Li, C. Liu, L. Liu, P. Sun, J.-J. Wu, Z. Xiao *et al.*, Lattice QCD study of the $K^*(892)$ resonance at the physical point, [arXiv:2603.16266](#).
- [54] Belle-II Collaboration, Measurement of inclusive $B \rightarrow X_u \ell \nu$ partial branching fractions and $|V_{ub}|$ at Belle II, *Phys. Rev. D* **113**, 032004 (2026).
- [55] A. Broggio, B. Capdevila, A. Ferroglia, and P. Gambino, Structure functions for the inclusive semileptonic b -quark decay at NNLO: A semi-analytic calculation, *J. High Energy Phys.* **06** (2026) 121.
- [56] L. Chen, X. Chen, X. Guan, and Y.-Q. Ma, Triple differential heavy-to-light semi-leptonic decays at next-to-next-to-next-to-leading order in QCD, [arXiv:2602.11537](#).
- [57] M. Black, R. V. Harlander, J. T. Kohnen, F. Lange, A. Rago, A. Shindler *et al.*, Bag parameters for heavy meson lifetimes, [arXiv:2603.28516](#).
- [58] M. Black, R. V. Harlander, J. T. Kohnen, F. Lange, A. Rago, A. Shindler *et al.*, Heavy-meson bag parameters using gradient flow, [arXiv:2603.28517](#).
- [59] Particle Data Group, Review of particle physics, *Phys. Rev. D* **110**, 030001 (2024).
- [60] Heavy Flavor Averaging Group (HFLAV), Averages of b -hadron, c -hadron, and τ -lepton properties as of 2023, *Phys. Rev. D* **113**, 012008 (2026).
- [61] F. Bernlochner, M. Jung, M. Khan, G. Landsberg, and Z. Ligeti, Novel approaches to determine $B^\pm$ and B^0 meson production fractions, *Phys. Rev. D* **110**, 014007 (2024).
- [62] M. Jung and S. Schacht, $\bar{B} \rightarrow D^{(*)} \ell \bar{\nu}$ branching ratios and evidence for isospin breaking in $\Upsilon(4S)$ decays, [arXiv:2604.08391](#).
- [63] Belle Collaboration, Measurement of the B^+/B^0 production ratio in e^+e^- collisions at the $\Upsilon(4S)$ resonance using $B \rightarrow J/\psi(\ell\ell)K$ decays at Belle, *Phys. Rev. D* **107**, L031102 (2023).
- [64] R. Fleischer and T. Mannel, General analysis of new physics in $B \rightarrow J/\psi K$, *Phys. Lett. B* **506**, 311 (2001).
- [65] LHCb Collaboration, Differential branching fractions and isospin asymmetries of $B \rightarrow K^{(*)} \mu^+ \mu^-$ decays, *J. High Energy Phys.* **06** (2014) 133.
- [66] LHCb Collaboration, Measurements of the S-wave fraction in $B^0 \rightarrow K^+ \pi^- \mu^+ \mu^-$ decays and the $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ differential branching fraction, *J. High Energy Phys.* **11** (2016) 047.
- [67] CMS Collaboration, Test of lepton flavor universality in $B^\pm \rightarrow K^\pm \mu^+ \mu^-$ and $B^\pm \rightarrow K^\pm e^+ e^-$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV, *Rep. Prog. Phys.* **87**, 077802 (2024).
- [68] Belle Collaboration, Observation of a new charged charmoniumlike state in $\bar{B}^0 \rightarrow J/\psi K^- \pi^+$ decays, *Phys. Rev. D* **90**, 112009 (2014).
- [69] LHCb Collaboration, First observations of the rare decays $B^+ \rightarrow K^+ \pi^+ \pi^- \mu^+ \mu^-$ and $B^+ \rightarrow \phi K^+ \mu^+ \mu^-$, *J. High Energy Phys.* **10** (2014) 064.